

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO ENGENHARIA GEOGRÁFICA, GEOFÍSICA E ENERGIA



Identification of Iberian large fires climate conditions

Inês dos Santos Vieira

Mestrado em Ciências Geofísicas
Especialização em Meteorologia e Oceanografia

Dissertação orientada por:
Professor Doutor Ricardo Machado Trigo
Doutora Ana Cristina Machado Russo

2019

Acknowledgements

Firstly, I want to thank my two advisors, Ana and professor Ricardo, for all their advice, availability, patience and support during this journey.

Thanks to all colleagues in the lab 8.3.15. In particular to Miguel for his help in the initial data processing and Python's tips, to Sílvia for being always available and to Virgílio for his advice during the writing of the dissertation.

A very important and special thank you to my parents for always accompanying and supporting me during all phases of this academic career. Thank you very much.

Finally, this work was partially supported by national funds through FCT (Fundação para a Ciência e a Tecnologia, Portugal) under project IMPECAF (PTDC/CTA-CLI/28902/2017).

Resumo

Os incêndios florestais são um risco que não é recente na Península Ibérica. No entanto, é inquestionável que alguns dos episódios mais dramáticos, tanto em termos de vítimas infligidas como de área total ardida, ocorreram durante este século, particularmente na região Oeste da Península Ibérica. A sua ocorrência é responsável, todos os anos, por uma grande quantidade de área ardida, traduzindo-se com frequência em impactos humanos e socioeconómicos significativos. Como tal, existe um reconhecimento crescente que a identificação de uma multiplicidade de padrões de circulação associados à ocorrência de fogos (Fire Weather Types, FWT) tem um interesse considerável em estudos de regimes de fogo, nomeadamente para 1) uma melhor compreensão da relação entre a meteorologia e a ocorrência de fogos; 2) explicar os padrões de severidade dos fogos; 3) melhorar as projeções/previsões e estratégias de combate aos fogos.

Neste estudo propõe-se a identificação e a análise dos fatores meteorológicos que controlam as variações da atividade associada aos grandes incêndios na Península Ibérica. Para isso, procede-se à classificação dos grandes de incêndios (aqueles cuja área ardida é superior ao percentil 95) para uma estação de fogos estendida (maio a outubro) para quatro regiões com regimes de fogo semelhantes. A identificação dessas regiões é feita de acordo com as condições meteorológicas locais (temperatura, humidade relativa, velocidade do vento e índices representativos da secura dos combustíveis), recorrendo-se a uma análise de compósitos para diferentes escalas temporais e de forma a captar a variabilidade interanual, subsazonal e sinótica. Para além disso, pretende-se ainda utilizar a análise de clusters (K-means) para identificar um conjunto limitado de FWT, sendo cada um caracterizado por uma certa combinação de condições meteorológicas sinérgicas condutoras do fogo.

A identificação dos limiares a partir dos quais um determinado fogo se considera um grande incêndio em cada uma das quatro regiões da Península Ibérica, permitiu a identificação de limiares distintos, sendo a região Este aquela que apresenta o valor mais elevado (171 ha).

A análise de compósitos aplicada a uma escala diária para as anomalias standardizadas das variáveis meteorológicas e para os índices que funcionam como proxies de seca, tendo em conta os percentis 1, 85, 95, 98 calculados para as áreas queimadas associados aos eventos ocorridos em cada uma das regiões, permite verificar um aumento das anomalias (em módulo) com o aumento da área queimada dos incêndios. A uma escala de 12 dias, as variáveis meteorológicas são aquelas que têm maior importância na distinção da atividade associada aos grandes incêndios, embora com desfazamentos (lags) temporais diferentes para as quatro regiões. A uma escala mais longa (oito meses), os índices que traduzem a secura dos combustíveis a diferentes camadas do solo são aqueles que têm maior importância.

A aplicação da análise de clusters permitiu identificar três FWT distintos para cada uma das regiões. O FWT_1 caracteriza-se por anomalias da velocidade do vento elevadas (acima de um desvio padrão). No caso do FWT_2, as anomalias das variáveis meteorológicas são bastante menores, situando-se abaixo de um desvio padrão. Por fim, o FWT_3 distingue-se por apresentar elevadas anomalias da temperatura e humidade relativa (acima de um desvio padrão). As quatro regiões apresentam, no entanto, resultados diferentes para cada um dos FWT identificados. Para além disso, verifica-se que o FWT_3 é aquele a que estão associadas as condições de secura dos combustíveis mais intensa para todas as regiões.

A uma escala de 12 dias, verifica-se que: 1) para o caso do FWT_1 a influência do vento está restrita ao dia do fogo; 2) para o FWT_2 as anomalias variam em torno de zero sem nenhum padrão definido; 3). O FWT_3 não apresenta uma influência restrita ao dia do fogo, mas a sua influência varia nas quatro regiões.

A análise do estado da secura dos combustíveis associados aos eventos caracterizados por cada um dos FWT permitiu identificar que o FWT_3 é aquele a que corresponde a uma situação em que os índices representativos da secura dos combustíveis apresentam anomalias mensais mais elevadas quando comparadas com as anomalias do FWT_1 e FWT_2 para as quatro regiões.

Esta análise permitiu identificar as condições climáticas associados à ocorrência dos grandes incêndios para quatro regiões da Península Ibérica tendo em conta as características meteorológicas das regiões, permitindo assim, diferenciar as condições meteorológicas e as diferentes escalas temporais entre a atividade que está na origem de todos os incêndios e aquela que está associada aos grandes incêndios.

Palavras - Chave: Grandes Incêndios, Península Ibérica, Análise de Compósitos, K-means, Fire Weather Types

Abstract

The Mediterranean region is characterized by the frequent occurrence of summer wildfires representing an environmental and socioeconomic burden. Some Mediterranean countries (or provinces) are particularly prone to Large Fires (LF), namely Portugal, Galicia, Greece, and southern France. On the other hand, the Mediterranean basin corresponds to a major hotspot of climate change, and anthropogenic warming is expected to increase the total burned area due to fires in Mediterranean Europe.

Here, we propose to classify summer large fires for four regions of Iberia (with similar fire regimes) according to their local-scale weather conditions (i. e. temperature, relative humidity, wind speed) and fire danger weather indices (Duff Moisture Code and Drought Code). The composite analysis was used to investigate the impact of local and regional climate drivers at different time scales, and to identify distinct climatologies associated with the occurrence of LF in Iberia for an extended fire season (May to October). A Principal Component Analysis (PCA) was applied to identify the variables with the highest variance explained for the fire day. Also, cluster analysis was used to identify a limited set of Fire Weather Types (FWT), each characterized by a combination of meteorological conditions leading to a better understanding of the relationship between meteorology and fire.

For each of the regions, three FWTs were identified with different characteristics. The FWT_1 is characterized by high anomalies (above one std) of zonal wind velocity. The FWT_2 presents anomalies of the meteorological variables within the average (bellow one std). Finally, the FWT_3 is categorized by high positive temperature anomalies (above one std) and strong negative relative humidity anomalies (bellow one std). The methodology followed allowed to objectively identify for different regions of Iberia multiple fire climatologies associated with the occurrence of LF.

Keywords: Large Fires, Iberian Peninsula, Composite Analysis, K-means, Fire Weather Types

Index

Acknowledgements	II
Resumo	III
Abstract	V
Table of figures	VII
List of tables	VIII
List of abbreviations and symbols	IX
1. Introduction	1
2. Data	3
2.1. Rural fires	3
2.2. Meteorological variables	4
2.3. Fuel moisture codes	5
3. Methodology	8
3.1. Large fires classification	8
3.2. Composite Analysis	9
3.3. Principal Component Analysis (PCA)	9
3.4. K – Means algorithm	11
4. Results and Discussion	13
4.1. Characterization of the study area	13
4.2. Large fires classification	14
4.3. Composite Analysis	16
4.3.1. Day of fire	16
4.3.2. Twelve days	19
4.3.3. Five Weeks	20
4.3.4. Eight months	22
4.4. Principal Component Analysis	23
4.5. K – Means Analysis: Fire Weather Types identification	24
4.6. Fire Weather Types characterization	27
4.6.1. Twelve days	27
4.6.2. Eight months	29
5. Conclusions	30
6. References	33
7. Annexes	37

Table of figures

Figure 2.1. - Geographical boundaries of the four spatially homogeneous regions of IP: Southwest (SW), North (N), East (E) and Northwest (NW).	5
Figure 2.2 – Schematic representation of the structure of the Fire Weather Index (FWI) system. Source: CWFIS.....	6
Figure 3.1 – Summary of the Principal Components Analysis Approach.....	11
Figure 4.1 - Monthly average values of temperature (°C), relative humidity (%), zonal wind (m/s), meridional wind (m/s) and precipitation (mm) for the four regions: Southwest (SW), North (N), East (E) and Northwest (NW).	13
Figure 4.2 - As in Fig. 4.1 but for DMC and DC.	14
Figure 4.3 - Percentage of burned area during the months of May, June, July, August, September and October for the four regions.	14
Figure 4.4 - Representation of the following representative percentiles: 1, 85, 95 and 98 for the four regions.	15
Figure 4.5 - Interannual variability of the number of large fires per year observed in the four considered regions.	16
Figure 4.6 – Composites of standardized anomalies of temperature, relative humidity (RH) and zonal velocity of the wind (U) of the day of fire for the SW, N, E and NW regions according to the percentiles 1 (gray), 85 (purple), 95 (orange), 98 (red) of the final burned areas.....	18
Figure 4.7 – As in Fig. 4.6 but for DMC and DC.....	18
Figure 4.8 - Composites of standardized anomalies of temperature, relative humidity and zonal wind velocity at 12 days for all fires (gray) and for large fires (red) for regions under study.	19
Figure 4.9 - As in Fig. 4.8 but for DMC and DC.	20
Figure 4.10 - Composites of standardized anomalies of temperature, relative humidity and zonal wind velocity at 5 weeks for all fires (gray) and for large fires (red) for the regions.	21
Figure 4.11 - As in Fig. 4.10 but for DMC and DC.	21
Figure 4.12 - Composites of the monthly standardized anomalies for eight months temperature, relative humidity and zonal wind velocity for all fires (grey) and for large fires (red).....	22
Figure 4.13 - As in Fig. 4.12 but for DMC and DC.	23
Figure 4.14 – Explained variance for the six principal components (PC 1 through PC 6) corresponding to the temperature (red), relative humidity (green), wind direction (blue), DMC (black), DC (pink) and precipitation (yellow) for the four regions of Iberia. The dashed blue line symbolizes the cumulative explained variance.	24
Figure 4.15 - Composites of the standardized anomalies of the meteorological variables (temperature, relative humidity and zonal wind) of the fire day associated with the three Fire Weather Types identified by K-means, FWT_1 (blue), FWT_2 (green) and FWT_3 (red).....	25
Figure 4.16 - As in Fig. 4.15 but for DMC and DC.	26
Figure 4.17 - Composites of the standardized anomalies of the meteorological variables (temperature, relative humidity and zonal wind) at 12 days timescale associated with the three Fire Weather Types identified by K-means, FWT_1 (blue), FWT_2 (green) and FWT_3 (red).....	28
Figure 4.18 – As in Fig. 4.17 but for DMC and DC.....	28
Figure 4.19 - Composites of the standardized anomalies of DC and DMC at 8 months timescale associated with the three FWT.	29
Figura A.1 – Trends of FWT_1 annual number of large fires (Mann- Kendall Test $p < 0.001$) for the SW, N and E with a negative trend and NW with no trend.....	37
Figure A.2 – Trends of FWT_2 annual number of large fires (Mann- Kendall Test $p < 0.001$) for the SW and NW with a negative trend and N and E regions with no trend.	38
Figure A.3 - Trends of FWT_3 annual number of large fires (Mann- Kendall Test $p < 0.001$) for the N and E with a negative trend and NW and SW regions with no trend.	39

List of tables

Table 2.1 - Monthly amounts of the empirical Daylength factor used to compute DMC and the Daylength Adjustment to calculate DC (Wagner, 1987).	7
Table 3.1 - Bibliographical references on the spatial resolution used in the studies on the importance of meteorological factors in the development of large fires, the type of databases for the fire event and also the definition of a large fire considered.....	8
Table 4.1 - Values of representative percentiles: 1, 85, 95 and 98 for the four regions.....	15
Table 4.2 - Representation of the relative importance of the Fire Weather Types identified in each region in terms of the percentage of Large Fires (LF) of the contribution of the respective burned area.	27

List of abbreviations and symbols

BA	Burned Areas
BUI	Buildup Index
CWFIS	Canadian Wildland Fire Information System
D	Present day's DC
DC	Drought Code
DMC	Duff Moisture Code
E	East Region
ECMWF	European Centre for Medium-Range Weather Forecasts
EFFIS	European Forest Fire Information System
EMS	Emergency Management Service
FFMC	Fine Fuel Moisture Code
FWI	Fire Weather Index
FWT	Fire Weather Types
ha	Hectares
ICNF	Institute of Nature Conservation and Forests
IP	Iberian Peninsula
IPCC	Intergovernmental Panel on Climate Change
ISI	Initial Spread Index
K	log drying rate in DMC
L_e	Daylength Factor in DMC
L_f	Daylength Adjustment in DC
LF	Large Fires
M	Moisture Content after rain in DMC
M_0	Moisture Content from previous day in DMC
MARM	Spanish Ministry of the Environment, and Rural and Marine Affairs
MSG	Meteosat Second Generation
N	North Region
NGP	Normalised Geostationary Projection
NW	Northwest Region
P	Precipitation
P_0	Previous day's DMC
PCA	Principal Components Analysis
PE	Potential Evapotranspiration
Q	Moisture Equivalent of DC
Q_0	Moisture Equivalent of previous day in DC
Q_r	Moisture Equivalent after rain
r_0	Precipitation of latest 24 hours
r_d	Effective Rainfall, DC
r_e	Effective Rainfall, DMC
RH	Relative Humidity
std	Standard Deviation
SVD	Singular Values Decomposition
SW	Southwest Region
T	Temperature
U	Zonal Wind Component
V	Meridional Wind Component
VLF	Very Large Fires

1. Introduction

Wildfires are a recurrent natural phenomenon (Bowman et al. 2009) which occurs with varying regularity and severity across almost every biome on Earth (Archibald et al. 2013). The Mediterranean region is characterised by the frequent occurrence of large summer wildfires and constitutes the region where most of the total burned area in Europe occurs, with an average of 4500 km² per year (Turco et al., 2017). However, the relevance of wildfires is not homogeneous within the Mediterranean basin, with some specific countries (or provinces) being particularly prone to Large Fires (LF), namely Portugal, Galicia (Spain), Greece, southern France (Trigo et al., 2013; Ruffault et al., 2016; Pinto et al., 2018). Unquestionably, some of the most dramatic episodes occurred during this century, particularly in the western region of the Iberian Peninsula (Sousa et al. 2015), often translating into significant human and socio-economic impacts (Trigo et al. 2013, Sánchez-Benítez et al., 2018). Numerous features make the landscape of Mediterranean Europe dissimilar from those of the rest of European continent, and these differences are mostly related to the climate and the long and intense impact of humans, but also reflect the role of fire (Pausas and Vallejo, 2011).

In the last decades, the European Mediterranean countries have experienced profound social and economic changes, in particular, the depopulation of rural areas, leading to the abandonment of large areas of cultivated areas which have led to the recovery of vegetation and an increase in accumulated fuel causing an increase in fire frequency (Pereira et al. 2005, Costa et al. 2011). Apart from the social and economic factors, weather and climate play a vital role as a distinguishing factor, so that most fires occurring in Europe occur in the Mediterranean basin. Moreover, most of the fires occur during the summer months, when the high temperatures and the low air humidity and fuel moisture increase the risk of fire (Trigo et al. 2006, Turco et al. 2013, Amraoui et al. 2015, Erickson et al. 2016). Furthermore, according to the latest report by IPCC (SR 1.5, 2018), it is expected an increase in mean global temperatures, including for the Mediterranean Europe area. However, whereas in the Mediterranean regions a decrease in precipitation is projected which has already started to occur (Ozturk al. 2015, Polade et al. 2017), leading to the occurrence of more frequent, prolonged and intense droughts in the Mediterranean, the remaining regions of Europe, especially in the North, an increase in precipitation is expected, and an increase in the occurrence of droughts is not anticipated (IPCC SR1.5, 2018). Besides these differences, the total burned area depends on a multiplicity of another factor, such as topography, biomass production and its availability to burn, ignitions patterns (Archibald et al. 2009).

The role played by climate and weather in LF occurrence has been frequently discussed (Flannigan et al., 2000; Pereira et al., 2005), putting in evidence that they both act at different time scales (Barbero et al., 2015). Ruffault et al. (2016, 2018) describe in details the role of meteorological and climatological variables in the occurrence of fires of different sizes in southern France, showing that at an annual to seasonal time scales, the atmospheric conditions are essential to the growth of the fine fuels. Besides, before and after the fire season, daily to monthly atmospheric conditions are responsible by the changes in moisture content, and finally, the instantaneous conditions (hourly to daily) in meteorological variables may influence fire ignition, intensity and propagation (Riley et al. 2013, Barbero et al. 2015).

The dependency of fire activity on meteorological conditions can be quantified using directly meteorological variables or using indices that combining several meteorological variables. Fire danger rating systems try to anticipate periods of heightened fire risk, mainly for early warning to the authorities. One of these systems, most likely the best known in Europe and North America, is the Canadian Forest Fire Weather Index System (FWI), developed by the Canadian Forestry Service, is

often used to rate the fire danger. The FWI is based on daily values of temperature, relative humidity, wind at noon and 24-hour accumulated precipitation, and has been operational since 1970 in Canada (Van Wagner, 1987) being the most extensive fire danger rating system used operationally in the world (Field et al., 2015).

In Europe, the FWI system is fully operational and is provided every day by the EFFIS (European Forest Fire Information System) which is part of the Copernicus Emergency Management Service (Copernicus EMS). The Copernicus EMS was developed so it “provides information for emergency response in relation to different types of disasters, including meteorological hazards, geophysical hazards, deliberate and accidental man-made disasters and other humanitarian disasters as well as prevention, preparedness, response and recovery activities”. The Copernicus EMS is composed of an on-demand mapping component providing rapid maps for emergency response and risk and recovery maps for prevention and planning and of the early warning and monitoring component which includes systems for floods, droughts and forest fires (Copernicus Emergency Management Service, 2019).

However, there is considerable interest in identifying the meteorological factors that control the variations in LF activity under a global change context, since, there are still few studies on the impacts of climate change on fire risk for the Iberian Peninsula. For example, Sousa et al. (2015) projected for future climate scenarios, an increase in mean burnt areas for the considered pyro-regions of the Iberian Peninsula, with an increase of burned areas that can be two to three times higher than the current one. Whereby there is growing recognition that the identification of a multiplicity of Fire Weather Types (FWT) has a considerable interest in studies of fire regimes, namely for (1) a better understanding of the relationship between meteorology and the occurrence of fires; 2) explain the patterns of fire severity; 3) to improve projections/forecasts and fire-fighting strategies at a regional scale (Ruffault et al. 2016).

This study focuses on the analysis of historical meteorological data and records of fires and it aims to classify large summer fires for four regions of Iberia according to their local-scale weather conditions (i. e. temperature, relative humidity, wind speed) and fire danger weather indices that are components of the Canadian FWI (duff moisture code and drought code). The composite analysis was used to investigate the impact of local and regional climate drivers at different time scales, and to identify distinct climatologies associated with the occurrence of LF in Iberia.

2. Data

2.1. Rural fires

Information on each fire event recorded between 1980 and 2015 is taken from two separate databases. For Portugal, it was used the information provided by the Institute of Nature Conservation and Forests (ICNF), and for Spain, the data was provided by the Ministry of Environment, Rural, and Marine Affairs (MARM) of the Government of Spain.

The ICNF rural fire database is not consistent throughout the study period, and the information provided is different (Pereira et al., 2011). Two different periods can be considered, the first from 1980 to 2000, which for each fire includes information about the district, municipality and the parish of the ignition, the date and time of the beginning and the extinction of the fire, the total burned area of the fire in hectares, which corresponds to the sum of the forested areas and non-wooded areas and, finally, the cause of the fire when known. After 2001, the great addition to the previous information is that the “exact” location of the beginning of the fire in XY coordinates was added instead of considering as previously that the fire began in the centre of the parish (Pereira et al., 2011).

The Spanish database is considered one of the oldest in Europe, providing fires data since 1968 (Velez, 2001), though the records are not considered entirely consistent until 1988. All the information included in this dataset is compiled from the information provided by the forest fire reports from the autonomous regions (Moreno et al., 2011). Accurate information about the fire location is only available since 2000, where the utilization of GPS improved the knowledge of fire ignitions. Here the fire ignition documented is based on a spatial reference grid of 10×10 km provided by ICONA, the firefighting services use this grid for the determination of the approximate location of fire events, and moreover, the municipality origin of the ignition is also recorded. The method described by De la Riva et al. (2004) to locate the start of the fires successively refines and decreases the potential location area of the ignition points by excluding areas where the fire could not have occurred. It starts in the 10×10 grid, i.e. with a potential location area of 100 km². Later this area is decreased by intersecting with the frontiers of the municipality origin of the fire. Lastly, the location area is restricted to the forest perimeter (MARM, 1997) to determine the final potential location area. The use of this technique leads to a substantially smaller area where the ignition points are then randomly distributed (Rodrigues and De la Riva, 2014).

For this study, it is considered a prolonged fire season is spanning between May and October, since, most of the large fires in the Iberian Peninsula occurs during the summer period. The analysis was also restricted to burned areas above 1 ha in order to exclude fires whose development was not forced by meteorological factors. Besides that, the fire events of both datasets are organised by date. Each event is also spatially organised by the 18 districts in Portugal and 48 provinces in Spain (Fig. 2.1).

Afterwards, each fire was associated to one of the four regions identified by Trigo et al. (2013) for the Iberian Peninsula (IP), where a cluster analysis allowed the identification of four spatially homogeneous regions (Southwest (SW), North (N), East (E) and Northwest (NW)) with a similar fire regime (Figure 2.1).

2.2. Meteorological variables

The meteorological variables used in this analysis are taken from the ERA-Interim archive (Dee et al., 2011). ERA-Interim is a global atmospheric reanalysis produced by the European Centre Medium-Range Weather Forecasts (ECMWF) and is provided with a sequential data assimilation scheme, advancing forward in time using 12-hourly analysis cycles, where the observations are combined with prior information from the forecast model to estimate the evolving state of the global atmosphere and its underlying surface (Dee et al. 2011). This reanalysis covers the period since 1979 and continues to be updated until 31 August 2019. The ERA-Interim project was developed with the purpose of preparing a future reanalysis covering the whole of the 20th century but the main objectives of the project were to improve some critical features of ERA-40, such as the representation of the hydrological cycle, the quality of the stratospheric circulation, and the handling of biases and changes in the observing system (Berrisford et al. 2011).

The spatial resolution used by ERA-Interim configured for the data set is approximately 79 km (reduced Gaussian grid) on 60 vertical levels from the surface up to 0.1 hPa. The ERA-Interim data assimilation and forecast suite produce four analyses per day, at 00, 06, 12 and 18 UTC and two 10-day forecasts per day, initialised from analyses at 00 and 12 UTC. For detailed documentation of the ERA-Interim archive, see Berrisford et al. (2011) and Dee et al. (2011).

For this study, five daily surface variables at 12 UTC were extracted from the ERA-Interim reanalysis database, namely, dewpoint temperature at 2 m, the temperature at 2 m, zonal wind component at 10 m, meridional wind component at 10 m and accumulated precipitation of the last 24 hours. In addition to these variables, the relative humidity was computed from the Magnus formula, depending on dewpoint temperature and temperature at the surface, described by Lawrence (2005).

The wind intensity was calculated from the module of the zonal and meridional wind components for each grid point.

The meteorological data were re-projected in the normalised geostationary projection (NGP) of Meteosat Second Generation (MSG) following Pinto et al. (2018) since the FWI database is calculated for the MSG disk, as well as the respective indexes that give rise to it, including DMC and DC. Afterwards, each grid point is attributed to one of the four regions identified by Trigo et al. (2013) for the IP (Figure 2.1).

The meteorological data were then put through a preliminary pre-processing sequence according to the following procedure:

1. Calculation of daily, weekly and monthly climatologies for the mentioned variables and using as reference period the years from 1980 to 2015.
2. Calculation of daily, weekly and monthly anomalies for all variables in the study.
3. Standardization of all variables to allow comparison between variables.



Figure 2.1. - Geographical boundaries of the four spatially homogeneous regions of IP: Southwest (SW), North (N), East (E) and Northwest (NW).

2.3. Fuel moisture codes

The FWI system is organised into six different components reflecting distinct aspects of the impact of meteorological variables in the fuels flammability and fire spreading characteristics (Figure 2.2). FWI system computation requires measurements of instantaneous temperature at 2 m, relative humidity at 2 m and sustained wind speed at 10 m, all at noon and precipitation totalled over the previous 24 h. Firstly the FWI system computes the fuel moisture codes that follow daily changes in the moisture contents of three different layers of forest fuel with varying rates of drying, in second place is performed the two intermediate components of the system which demonstrating rate of spread and amount of available fuel. Moreover, finally, it is calculated the Fire Weather Index (Figure 2.2). The FWI is an amalgamation of the ISI (Initial Spread Index) that quantifies the rate of spread alone without the influence of variable quantities of fuel, and the BUI (Buildup Index) that accounts for the total fuel available to the spreading fire, indicating the intensity of the spreading fire as energy output rate per unit length of the fire front (Wagner, 1987).

In this study, two codes associated to the BUI calculation were considered in further detail, namely the Duff Moisture Code (DMC) and the Drought Code (DC). Both the DMC and DC are numerical values that represent the average moisture content of fuels. In case of the DMC, of the loosely compacted organic layers of moderate depth equivalent, a duff layers about 7 cm deep and 5 kg/m² in dry weight and the primary goal of this code is indicating fuel consumption in moderate duff layers and medium-size woody material. The DMC is computed through the following equation (Van Wagner, 1987):

$$DMC = P + 100K \quad (2.1)$$

DMC is computed following a two-stage process. The first one is the rainfall phase, which aims to calculate the actual moisture content that rises with dryness through the following expression:

$$P = 244.72 - 43.43 \ln(M - 20) \quad (2.2)$$

The computation of the moisture content after rain (M) that is a function of effective rain (r_e), a coefficient b which is performed by one of a set of three empirical equations depending on the value of the initial code (P_0), and initial moisture content (M_0):

$$M = M_0 + 1000 r_e / (48.77 + b r_e) \quad (2.3)$$

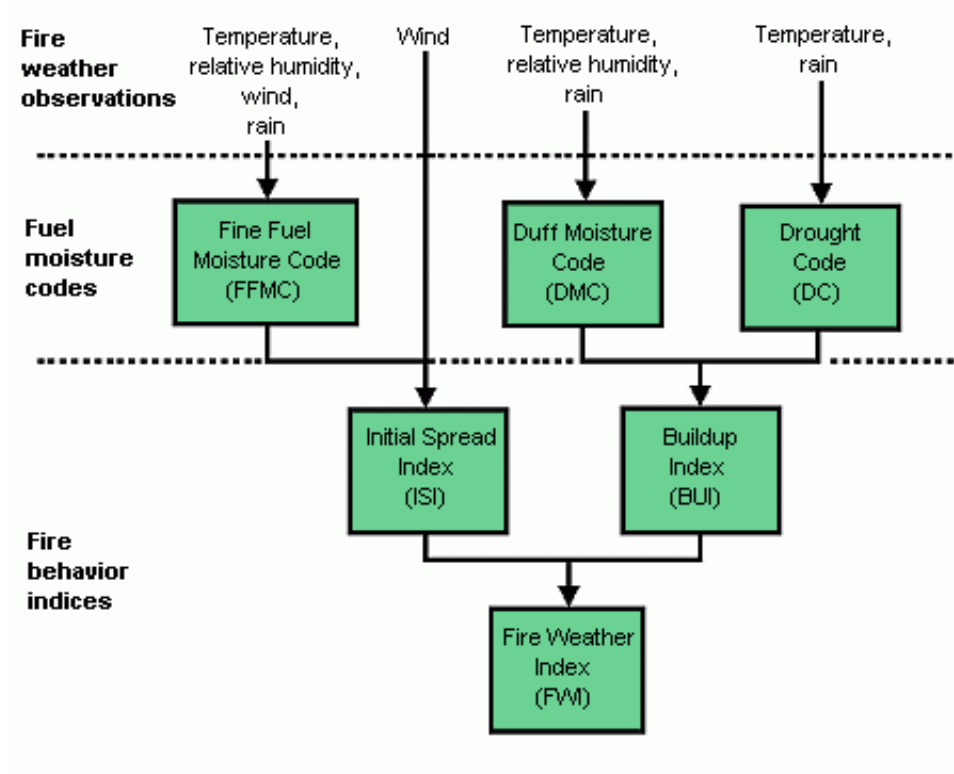


Figure 2.2 – Schematic representation of the structure of the Fire Weather Index (FWI) system. Source: CWFIS.

Subsequently, to perform the drying phase, we must compute the log of the drying rate K (given by the following expression) at noon, which is a function of temperature (T) and relative humidity (RH), and also an empirical daylength factor (L_e) listed by month in Table 2.1, which is given by the equation:

$$K = 1.894 (T + 1.1)(100 - RH)L_e 10^{-6} \quad (2.4)$$

The DC, on the other hand, represents the average moisture content of deep, compacted organic layers and heavy surface fuels, about in average 18 cm deep and 25 kg/m² in dry weight. DC is calculated through the following equation Van Wagner, 1987):

$$DC = D + 0.5PE \quad (2.5)$$

Where D represents the current drought, and PE corresponds to potential evapotranspiration. As in the DMC, the DC computation involves the calculation of the rainfall phase and the drying phase. The calculation begins with a rainfall phase, which represents the moisture equivalent after rain, which is in the function of the effective rainfall (r_d) and the existing moisture equivalent, given by the following expression:

$$Q_r = Q_0 + 3.937r_d \quad (2.6)$$

$$r_d = 0.83r_0 - 1.27 \quad (2.7)$$

If the precipitation (r_0) of the latest 24 hours it is smaller than 2.8 mm, the component of the rainfall is ignored for the DC calculation. Therefore, we can perform the calculation of the numerical value of the current drought through the next equation:

$$D = 400 \ln(800/Q) \quad (2.8)$$

The drying phase, which is represented through an empirical expression, evaluates the potential evapotranspiration (PE) depending on the temperature (T) at noon and the seasonal Daylength adjustment (L_f) – values at table 2.1:

$$PE = 0.36(T + 2.8) + L_f \quad (2.9)$$

Table 2.1 - Monthly amounts of the empirical Daylength factor used to compute DMC and the Daylength Adjustment to calculate DC (Wagner, 1987).

MONTH	EMPIRICAL DAYLENGTH FACTOR	DAYLENGTH ADJUSTMENT
JANUARY	6.5	-1.6
FEBRUARY	7.5	-1.6
MARCH	9.0	-1.6
APRIL	12.8	0.9
MAY	13.9	3.8
JUNE	13.9	5.8
JULY	12.4	6.4
AUGUST	10.9	5.0
SEPTEMBER	9.4	2.4
OCTOBER	8.0	0.4
NOVEMBER	7.0	-1.6
DECEMBER	6.0	-1.6

The DMC and the DC were both computed using ERA-Interim reanalysis described in the previous section. Moreover, both DMC and DC followed the same procedure described in the previous section for the meteorological variables where it is performed the climatology for the described period, the anomalies (diary, weekly and monthly) calculation and respective standardization of the anomalies to the four regions in the study.

3. Methodology

3.1. Large fires classification

There is an extensive bibliography on the meteorological factors that control the activity of large fires (e.g. Vázquez, Moreno J.M. 1995, Kasischke et al. 2002, Bradstock et al. 2009). However, there is no unique definition from which a fire is considered a large fire and different studies use distinct definitions (Table 3.1). While some studies follow official national definitions (e.g. Parente et al. (2016) follows the ICNF definition of a LF if the burned area is higher than 100 ha), other studies use statistical definitions (e.g. Pereira et al., 2005; Barbero et al., 2014, 2015). These later consider all those fires with burned areas higher than the 90th percentile of the respective distribution as an LF. Ruffault et al. (2016, 2018) use an even higher threshold, considering that fire is a LF if the burned area is above the 95th percentile of the respective distribution. The use of a percentile based definitions at the national level may not be the correct one, since, a range of factors influences the incidence and size of fires, besides the weather and climate, such as ignition sources, fuels, terrain and suppression forces (Ganteaume et al., 2013). Therefore, the use of a statistical definition should be used for piroregions where fire behaviour behaves similarly and not based on the borders of a country.

Therefore, following the definition proposed by Ruffault et al. (2016), each fire event was classified into four classes according to the final burned area. Each class is defined by the calculation of percentiles 1, 85, 95 and 98. Thus, within the scope of this work, a fire was considered a large fire if the burned areas, express in hectares (ha), presented a value higher than of the 95th percentile of the respective distribution, which was determined based on all dataset from 1980 to 2015. This definition is applied to the fire events attributed to the four regions under study (Fig 2.1), previously identified with similar fire regimes. In general, a fire regime characterises the spatial and temporal patterns and ecosystem impacts of fire on the landscape. Accordingly, these regions may be distinguished by how often fires typically occur (frequency, fire interval, fire rotation) and vegetation type (or ecosystem), in consequence of similar weather and climate patterns. As a result, it is estimated that the threshold from which one considers an LF in each of the regions has a different value.

Table 3.1 - Bibliographical references on the spatial resolution used in the studies on the importance of meteorological factors in the development of large fires, the type of databases for the fire event and also the definition of a large fire considered.

Reference	Spatial Resolution	Fire Datasets	LF Definition
Pereira M.G. et al. (2005)	National	In Situ information	> 90 th percentile
Ganteaume and Jappiot (2013)	Regional	In Situ information	> 100 ha
Barbero et al. (2014, 2015)	Regional	Satellite information	> 90 th percentile
Urbieta R.I. et al. (2015)	Regional	In Situ information	> 100 ha
Parente J. et al. (2016)	Regional	In Situ information	> 100 ha
Ruffault et al. (2016, 2018)	Regional	In Situ information	> 95 th percentile

3.2. Composite Analysis

One of the objectives of this study is the identification and analysis of the meteorological factors that control the variations of the LF activity that occur in the four Iberian regions (Figure 2.1). For this purpose, a similar approach to Ruffault et al. (2016, 2018) was used, i.e. a composite analysis for different temporal scales: twelve days, five weeks and eight months, in order to capture the synoptic, subsazonal and interannual variability, respectively.

After a preliminary treatment of the variables (described at section 2.2, 2.3), and considering the thresholds identified by the calculation of the percentiles of the burned areas for each area, two different subsets were considered for each region of the IP: i) for all the fires and ii) the fires with areas burned above the 95th percentile according to the different periodicities to analyse:

Daily: The daily composites of the variables were performed on a time scale of up to twelve days. The composites are computed based on the mean of the anomalies for several days. Firstly, the days of the ignitions are identified, and after identifying the day of the beginning of the fires is associated with the anomaly of temperature, relative humidity, zonal wind velocity, DMC and DC. Next, the mean of the anomalies for the identified fire days is calculated. This procedure was adapted for the remaining days, where after identifying itself, each of the days of the fire is calculated the mean of the anomalies for each of the days up to 8 days before and in the two days after the day of the beginning of the fire.

Weekly: The weekly composites of the variables are carried out on a time scale of five weeks. The same procedure was followed as for daily composites, but in this case, after identification of the fire day the climatological anomaly of each of the variables of the week in which the fire occurs is calculated and the weekly average of every week up to 4 weeks before the event and a week later.

Monthly: The monthly composites of the variables are performed on a time scale of eight months. Once again, after identifying the fire day, the average of the anomalies of the study variables associated with these days was calculated, but in this case, the monthly anomalies associated with each event was considered, and the average of the monthly anomalies are compiled up to 7 months before the fire month and a month after.

3.3. Principal Component Analysis (PCA)

Principal component analysis (PCA) is a statistical procedure where its main objective is to reduce the dimensionality of a dataset consisting of many interrelated variables while retaining as much as possible of the variation present in the data set (Jolliffe, 2002). This is achieved by performing an orthogonal transformation to obtain a new set of linearly uncorrelated variables, the principal components, and which are ordered so that the first few retain most of the variation present in all the original variables. Figure 3.1 is a schematic summary of the PCA approach.

The PCA method is widely used in many scientific fields of study such as in agriculture (Nikkhah et al. 2019), chemistry (Malik et al. 2018), climatology (João A. Santos and Margarida Belo-Pereira, 2019), ecology (Ciuti S. et al. 2018), economics (Fatoki Olawale and David Garwe, 2010), geology (Unglert et al. 2016), meteorology (Peña-Gallardo M. et al. 2019) and in many other areas.

Here, the PCA was implemented using Python *Scikit-Learn* library (F. Pedregosa et al., 2011), and it was used in order to identify the meteorological variables with higher explained variance at the start of the fire. In other words, PCA algorithm allows converting high dimensional data, which in this case includes temperature, relative humidity, zonal wind, duff moisture code, drought code and precipitation, to a lower dimensional data by selecting the most critical features that capture maximum information about the dataset in the day of the fire.

First, and since PCA yields a feature subspace that maximises the variance along the axes, it makes sense to standardise the data, especially, as in this case, if the different meteorological fields considered use widely different units and range of values. This procedure prevents that principal component from being biased towards features with high variance, leading to false results (Wilks, 2011). After preparing the dataset, PCA followed three main steps (Jolliffe, 2002):

1. **Identification of the eigenvectors/eigenvalues:** The foundation of the PCA is represented by the eigenvectors and eigenvalues of the covariance matrix or correlation matrix. The eigenvectors, also called principal components, determine the directions of the new feature space, and the eigenvalues determine their magnitude. That is, the eigenvalues explain the variance of the data along the new feature axes. Although eigen-decomposition of the covariance or correlation matrix may be more intuitive, the majority of PCA algorithms are based in a Singular Vector Decomposition (SVD) to improve the computational efficiency (M. Navas et al. 2009). The *Scikit-Learn* library that contains the function to perform the principal components uses the Singular Value Decomposition, and, in this case, with was performed the ‘full’ algorithm, where the exact full SVD is computed.
2. **Reduction of dimensionality:** As mentioned earlier, PCA is generally used to reduce the dimensionality of a dataset of the original feature space by projecting it onto a smaller subspace, where the eigenvectors will form the axes. However, the eigenvectors only define the directions of the new axis, to decide which eigenvectors can be dropped without losing too many information for the construction of lower-dimensional subspace, it is also necessary to know the corresponding eigenvalues. The lowest eigenvalues bear the minimum information about the distribution of the data, and those are the ones that can be dropped.
3. **Calculation of explained variance:** The algorithms used to perform PCA calculates a useful measure called "explained variance," which can be calculated from the eigenvalues. The explained variance tells us how much information can be attributed to each of the principal components. After deciding how much information we want to keep we finally passed to the construction of the projection matrix (matrix of our concatenated top k eigenvectors) that will be used to transform the data set onto the new feature subspace.

After the PCA algorithm was applied, the explained variance of each of the main components corresponding to each of the variables was calculated. The variables with the highest percentage of explained variance were selected, and together, they explained at least 80% of the total variance.

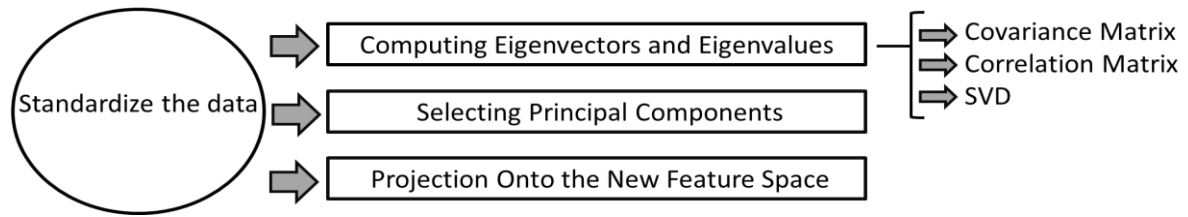


Figure 3.1 – Summary of the Principal Components Analysis Approach.

3.4. K – Means algorithm

The data clustering, also called as cluster analysis, is an essential tool of unsupervised learning and is defined in *Introduction to Data Mining and its applications* (2006), where it is described that the main objective of cluster analysis is to verify if the objects within a group are similar (or related) to one another and different from (or unrelated to) the objects in other groups. The higher the similarity (or homogeneity) within a group and the higher the difference between groups, the better or more distinct the clustering.

However, the concept of a group or cluster does not have a unique and precise definition. Thus, there is a wide range of methods and algorithms to perform cluster analysis, all of them with different procedures to define a cluster. Generally, clustering algorithms can be divided into three main groups, namely hierarchical, Bayesian and partitional algorithms (Saraswath S. and Sheela I. M, 2014). Hierarchical clustering algorithms recursively find clusters in an agglomerative mode, in other words, an algorithm starts by consider that each data point is a cluster and merge the most similar pair of groups successively to form a cluster hierarchy or in divisive mode, where all the data points is in one cluster and recursively dividing each cluster into smaller clusters. Bayesian algorithms follow a completely different approach, as they try to generate a posteriori distribution over the collection of all partitions of the data. Finally, partitional clustering algorithms find all the clusters simultaneously decomposing a data set into a set of disjoint clusters where the algorithms try to minimize an objective function (Jin X., Han J. 2011). One of the most popular and straightforward partitional algorithms is the K-means algorithm, which has been used across a broad range of application areas in many different fields. In fact, in the fields of meteorology and climatology, it has been successfully used by several authors, namely by Trigo I. F. et al. (1999), Skinner. et al. (2002) and Kassomenos P. (2010).

In the study, the K-means algorithm was implemented using Python Scikit-Learn library (F. Pedregosa et al. 2011) with the module `sklearn.cluster.Kmeans`.

The K-means algorithm clusters data by trying to separate samples in K groups of equal variances, such that the squared error between the empirical mean of a cluster and the points in the cluster is minimized (minimising the inertia or within-of-squares). The main goal of K-means algorithm is to reduce the sum of the squared error over all K clusters (inertia), where μ_k is the mean of cluster c_k , also called the within-cluster sum-of-squares criterion (A.K. Jain, 2010):

$$\sum_{k=1}^K \sum_{x_i \in c_k} (\|x_i - \mu_k\|)^2 \quad (3.1)$$

This expression describes a measure of how internally coherent clusters are. Where $X = \{x_i\}$, $i = 1, \dots, n$ correspond to dataset of n -dimensional points which the goal is to be clustered into K clusters, $C = \{c_k, k = 1, \dots, K\}$.

The crucial steps of the K-means algorithm are as follow (Tan et al. 2005):

1. Choose K initial points as centroids, with the most basic method being to choose K samples from the dataset X , where K is the number of clusters.
2. Each point is allocated to the nearest centroid.
3. After, the algorithm generates new centroids by taking the mean value of all the samples allocated to each previous centroid.
4. The difference between the last and the new centroids are computed, and the process is continued while this difference remains below a certain threshold defined by the user.

The initialisation of the K-means algorithm on Python language requires that the user specifies (scikit-learn user guide, 2019): the number of clusters, method of initialisation, the number of times the K-means algorithm will be run with different centroid seeds, the maximum number of iterations, relative tolerance to declare convergence and the algorithm to use. Nevertheless, one of the significant challenges in cluster analysis is the estimation of the optimal number of clusters. In this study, the ‘gap’ method was used (Tibshirani et al., 2001) before applying the cluster analysis to estimate the optimum number of clusters.

The K-means algorithm implemented, based on the results of the gap statistics method, was defined by the number of clusters for each region. Besides that, ‘K-means ++’ was used as the initialisation method that smartly selects initial cluster centres to speed up convergence. K-means will always converge, but different initialisations can lead to distinct final clustering because K-means can converge to a local minimum. Nevertheless, in this case, the K-means algorithm was run for 500 times with varying seeds of the centroid. Besides that, the relative convergence considered was the usual one, 0.0001, to declare convergence and for a single run, the maximum number of iterations considered was 300.

In this study, to reduce the dimensionality of the problem, the Principal Component Analysis method described in the previous section was applied. Afterwards, a cluster analysis algorithm based on the K-means method was applied to identify the FWT from standardized local scale anomalies associated with the LF previously identified for each of the four Iberian regions.

4. Results and Discussion

4.1. Characterization of the study area

The seasonal variability of the main meteorological variables considered for the period under study (1980-2015) can be observed in Figure 4.1 for the four regions of Iberia.

As expected, mean temperature (relative humidity and precipitation) values follow an annual cycle characterised by higher (lower) values during summer months. Usually, the SW and E regions present similar values in all the referred variables, whereas the NW is more similar to the N region. The zonal and meridional components of the wind present different behaviours for the four regions. The annual cycle of precipitation shows a well defined wettest season. In the winter months, the northernmost regions (NW and N) stand out with higher mean precipitation values.

In the case of DMC and DC (Figure 4.2), as in the case of meteorological variables, two groups of regions can be highlighted. The SW and E regions stand out with higher mean values of DMC and DC when compared to N and NW values. Moreover, the highest DMC (DC) values occur in August (September) in all four regions.

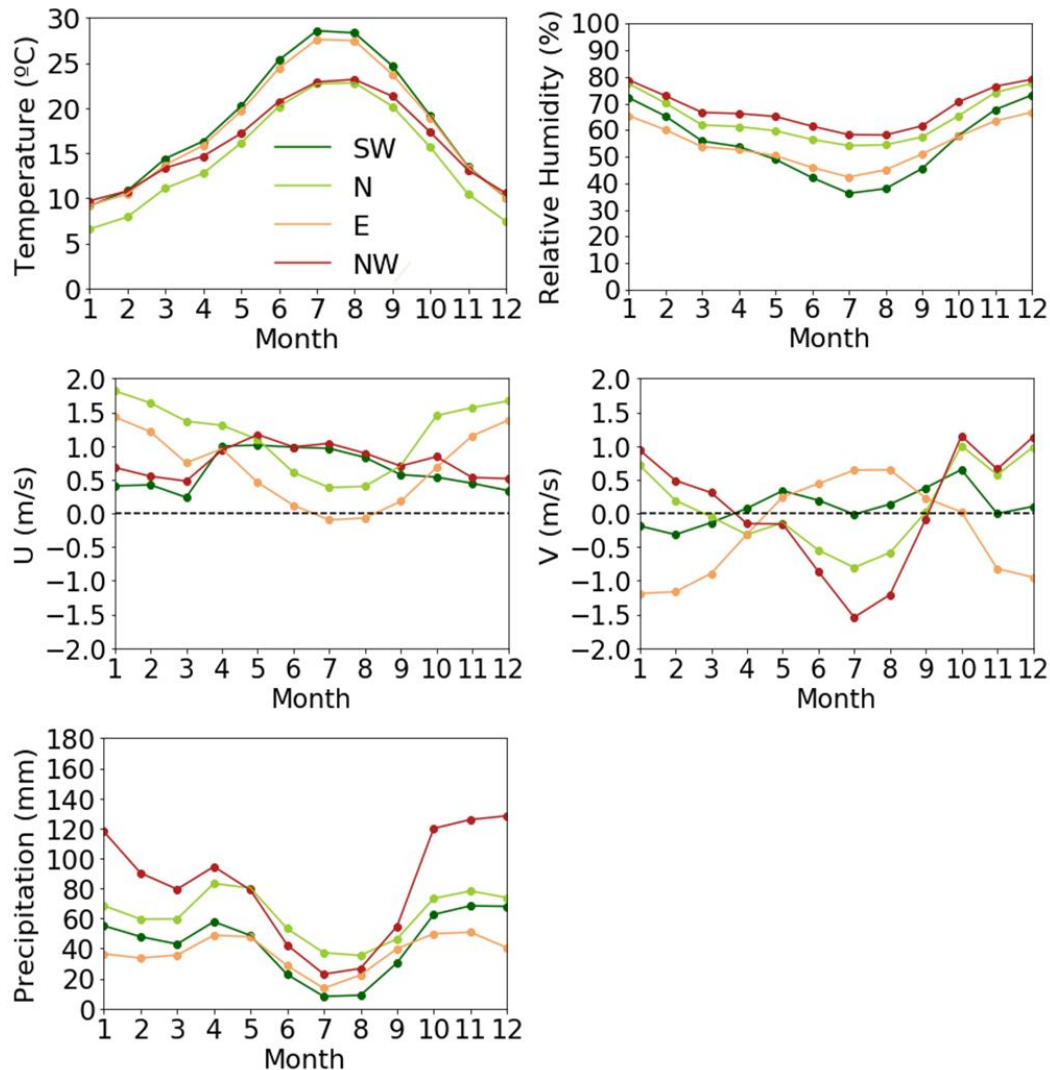


Figure 4.1 - Monthly average values of temperature (°C), relative humidity (%), zonal wind (m/s), meridional wind (m/s) and precipitation (mm) for the four regions: Southwest (SW), North (N), East (E) and Northwest (NW).

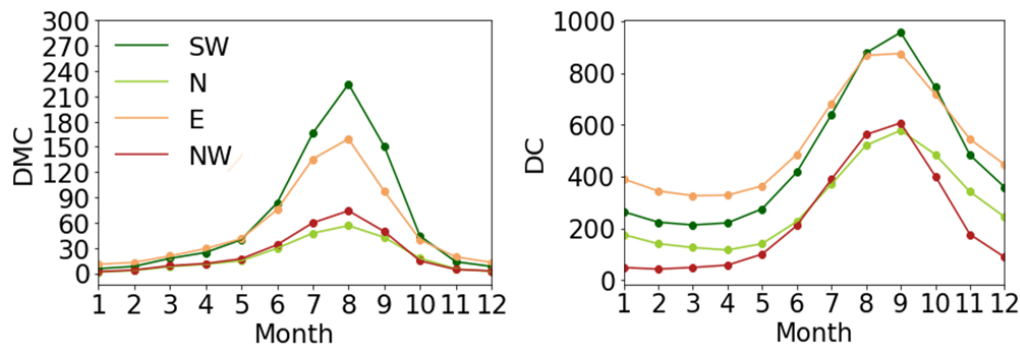


Figure 4.2 - As in Fig. 4.1 but for DMC and DC.

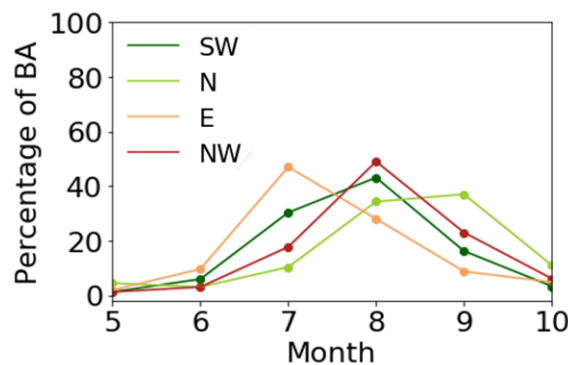


Figure 4.3 - Percentage of burned area during the months of May, June, July, August, September and October for the four regions.

The choice of these four regions of the Iberian Peninsula confirms the distinct spatial-temporal clusters that characterize them in the first place (Trigo et al., 2013). In fact, it is possible to observe that the relative amount of burned area associated with each of the regions differs, as well as the peak time month, with the region E presenting its maximum of BA (Burned Area) in July, while in the N region that occurs in September and the SW and NW clusters, that include all the Portuguese territory, present a maximum in August.

4.2. Large fires classification

As described in the methodology, all events with burned areas above the 95th percentile of the total distribution are considered as LF. Figure 4.4 shows the four histograms where the bars represent the 1st, 85th, 95th and 98th percentile, respectively, for each of the described regions with the same colours used in Figure 2.1 (this approach will be followed in all the remaining figures). In addition to figure 4.4, and in order to facilitate analysis, the corresponding percentile values are shown in table 4.1. The analysis allows us to verify that the thresholds of the BA are quite different among the regions. The exception being the 1st percentile, since the BA are filtered to consider the fires with final burned areas higher than 1 ha, the 1st percentile is 1 ha for all regions. The 85th percentile, although presenting different values for the four regions, is not drastically different among different regions. The E region presents the highest value (~32 ha), and the N region presents the lowest value (~20 ha). However, if

we consider the 95th percentile the threshold from which an LF is considered, the differences between regions are already considerable. The E region presents the highest burned area value of the four regions with 171 ha. The SW region is the one with the second highest threshold for an event to be considered an LF with about 125 ha there followed by the N region with 100 ha. Finally, the NW region is the one with the lowest threshold (~78 ha).

The very large fires (VLF) are characterised by having burnt area values above the 98th percentile threshold. Again, taking into account the results from Table 4.1, the E region is the one presenting the highest value of the burned area with 538 ha. In turn, the NW region has a threshold for which a VLF is considered with 200 ha.

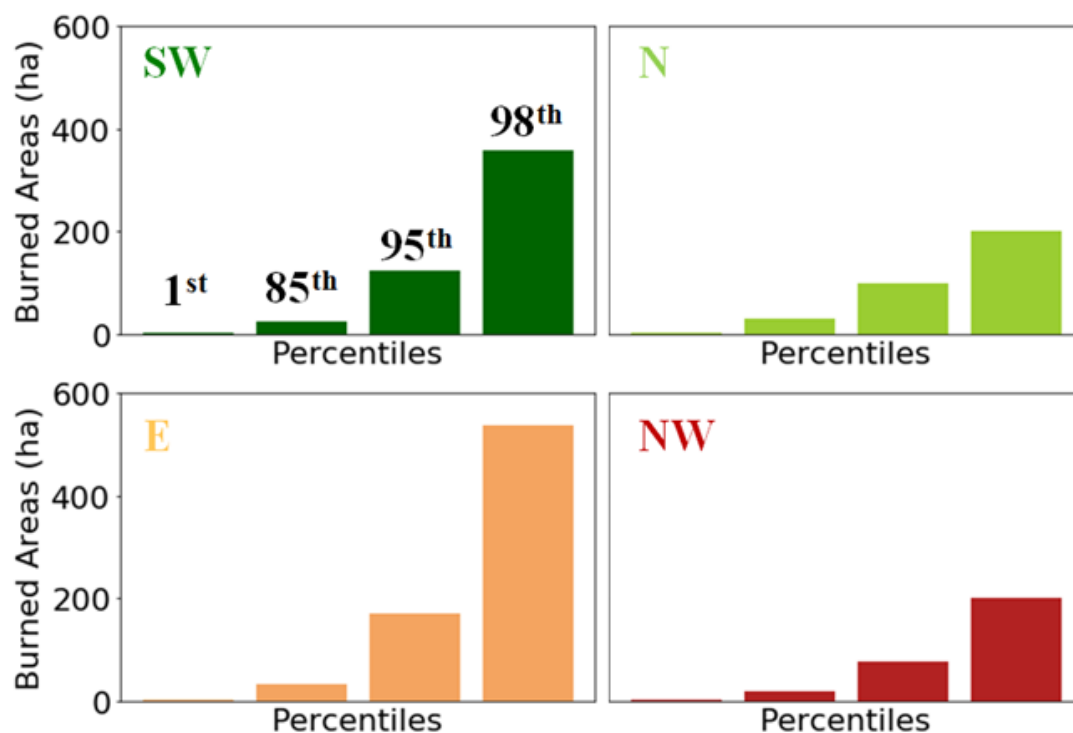


Figure 4.4 - Representation of the following representative percentiles: 1, 85, 95 and 98 for the four regions.

Table 4.1 - Values of representative percentiles: 1, 85, 95 and 98 for the four regions

Region	1 st (ha)	85 th (ha)	95 th (ha)	98 th (ha)
SW	1	25	125	357
N	1	30	100	202
E	1	32	171	538
NW	1	20	78	200

There is wide variability in the number of LF for the four regions of the IP (Figure 4.5). The NW region, although presenting the lowest LF threshold, is the one with the highest LF number and with the most fires (above 1 ha). The year of 1989 stands out with almost a thousand events. On the other hand, the E region, which presents the ultimate LF threshold of the four regions, is the one with the smallest LF number (below one hundred annual events).

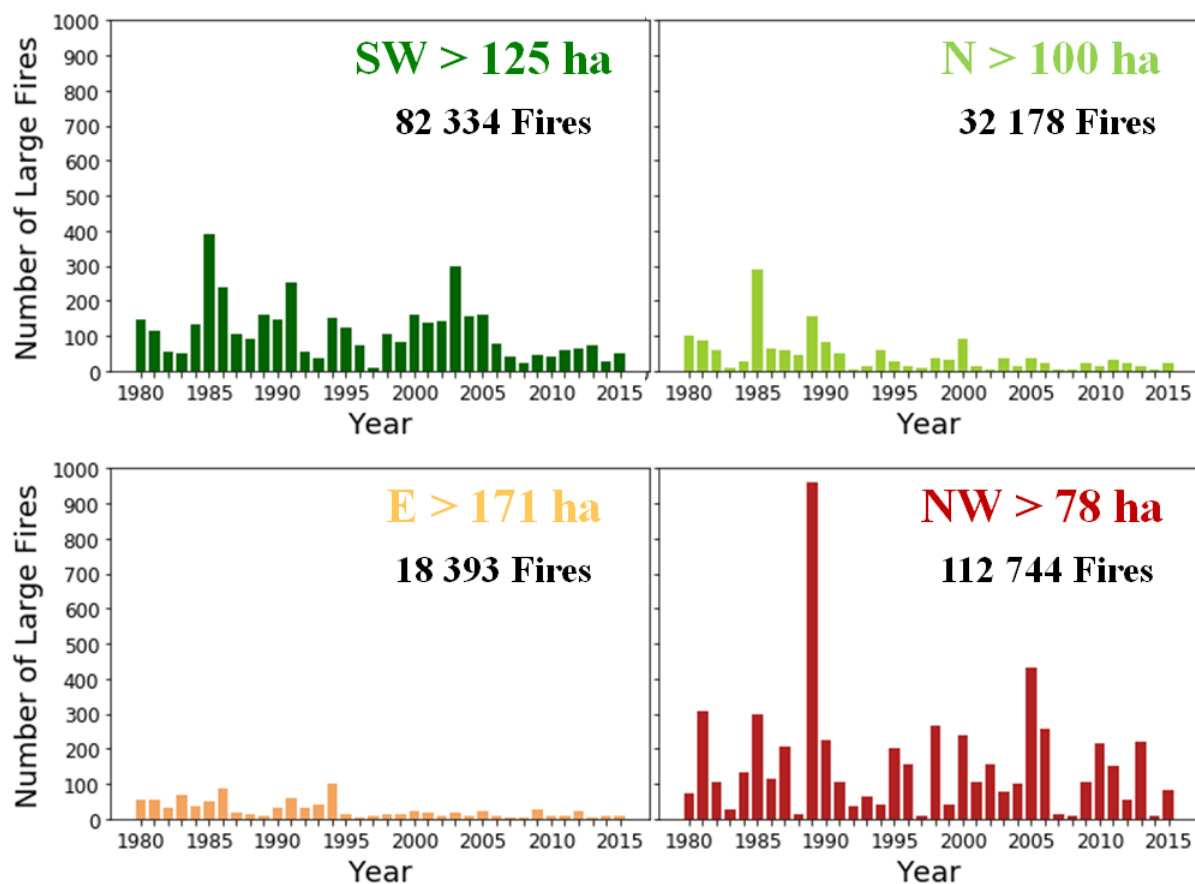


Figure 4.5 - Interannual variability of the number of large fires per year observed in the four considered regions and the total number of all fires (> 1 ha).

4.3. Composite Analysis

4.3.1. Day of fire

The standardized anomalies of the meteorological variables (temperature, relative humidity and zonal wind) for the day of the ignitions are recorded according to the percentile 1, 85, 95 and 98 of the total burned area for the four regions are represented in Figure 4.6. It is verified that for all regions, the anomalies of the meteorological variables increase in modulus according to the increase of the value of the total BA of the events. For the case of temperature, the standardized anomalies of the fire day are always positive and follow the behaviour already described, i.e., fires with a higher burned area are associated with days with higher temperature anomalies. In addition to this generalized behaviour, there are differences in the value of the anomalies when we compare results among the four considered regions. The NW region is the one that stands out with the highest anomalies of the meteorological variables for all classes of burned areas. At the opposite extreme lies the E region, where fires occurring in this region exhibit the lowest temperature anomalies associated with fire days, although the distinction

between all fires and large fires is visible. On the other hand, the SW and N regions have similar behaviour in the case of the temperature anomalies for LF and the VLF, although, in the case of the N region, the difference between LF and VLF is not significant. When we consider all the fires indiscriminately, the N region stands out with the highest temperature anomalies (1st percentile).

Relative humidity follows the opposite behaviour of temperature, as RH anomalies are increasingly negative for fires with higher burnt areas since its calculation depends directly on this variable.

Conversely, anomalies of zonal wind velocity present two very distinctive behaviours in the four regions, that accommodate the distinct signal (positive or negative) of zonal wind anomalies (U). Generally speaking the SW, N and NW regions present negative zonal wind (U) anomalies on the fire days, with this relationship getting stronger with the fire size. For these three regions, the negative anomalies are indicative of prevailing winds from the eastern quadrant. Moreover, once again, the NW region presents anomalies, in this case of U, more intense when compared with the other regions, and the N region does not present significant differences between the U anomalies associated with the LF and the VLF. On the contrary, for the E region, the U anomalies grow positively with the increase of the burned area associated with an event. The positive sign of U anomalies is indicative of prevailing winds from the western quadrant.

A similar analysis was made for the FWI indices (Figure 4.7), displaying the standardized anomalies of the indices of fuel dryness (DMC and DC). As in the case of the majority of the meteorological variables, DMC and DC present higher anomalies for higher burned areas. For the DMC, and contrary to the anomalies of the meteorological variables, it is the N region that presents the highest anomalies, and both, LF and VLF have DMC anomalies higher than one standard deviation. There is no difference between these two classes of burned areas and the same also occurs for the NW region, but with lower DMC anomaly values. Again, the E region is the one presenting the lowest anomalies when compared to the other regions, and it is possible to distinguish between the four classes of agreement of the final burned area, as in the SW region.

The magnitude of DC anomalies differs from the DMC anomalies. Both the NW and the N regions reveal the highest values of DC anomalies associated with the ignition days. In these two regions, there are variations in the value of DC anomalies for the four percentiles. Nevertheless, for the SW region, there are no differences in the anomalies between the four classes of burned areas.

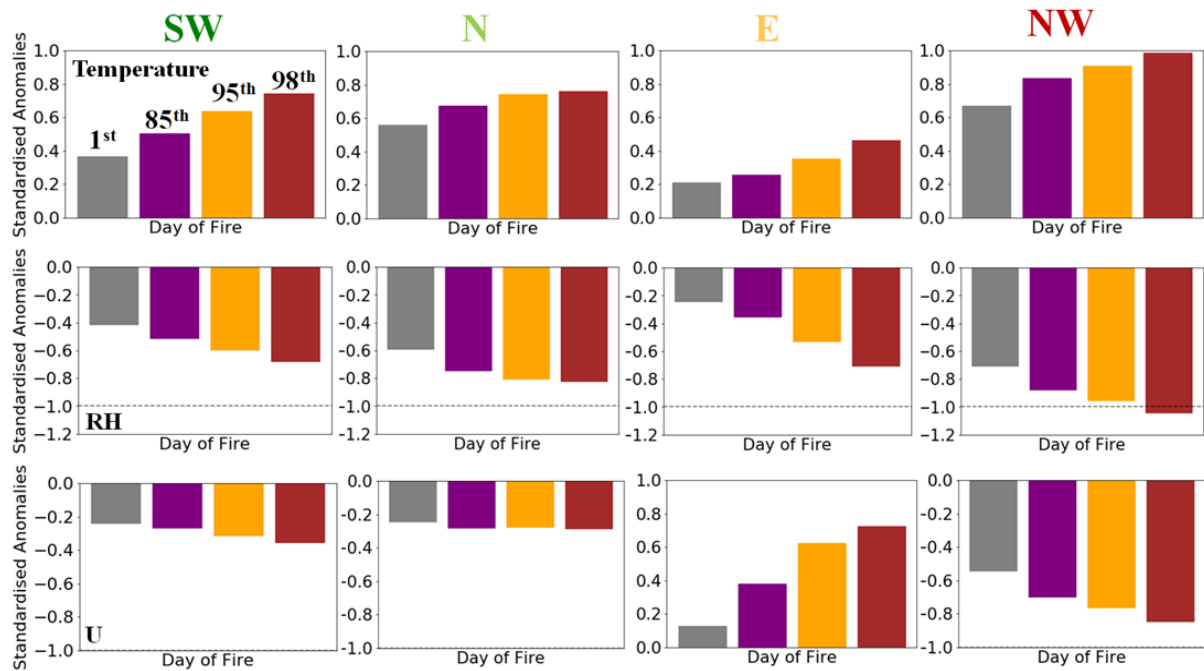


Figure 4.6 – Composites of standardized anomalies of temperature, relative humidity (RH) and zonal velocity of the wind (U) of the day of fire for the SW, N, E and NW regions according to the percentiles 1 (gray), 85 (purple), 95 (orange), 98 (red) of the final burned areas.

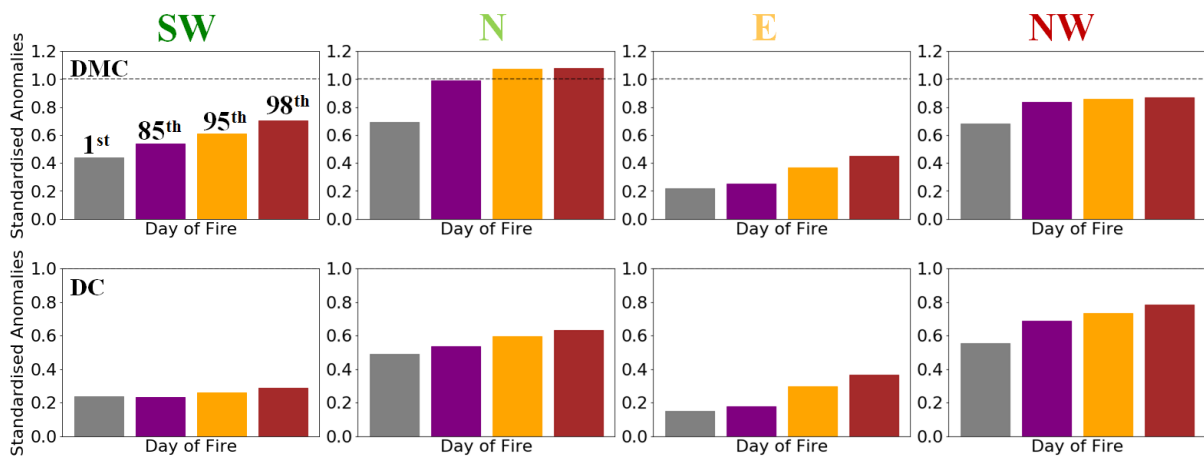


Figure 4.7 – As in Fig. 4.6 but for DMC and DC.

4.3.2. Twelve days

The daily standardized anomalies at a 12-day time scale (ten days before the start of the fire and two days after the beginning of the fire) of the temperature, relative humidity and zonal wind velocity are described on figure 4.8. With the representation of the composites for all the fires (grey) and the LF (red), it is intended to evaluate the influence of each of the meteorological variables on the activity of the fires and how the activity associated with the LF differs from the one associated to all fires activity. The zonal wind has a restricted influence on the day of the fire itself. In all cases, the anomalies decrease immediately after the fire day.

Looking at the SW region, it is possible to observe that both temperature and RH anomalies are higher (in absolute value) for the day of ignition and the previous day, particularly for LF. At the analyzed scale

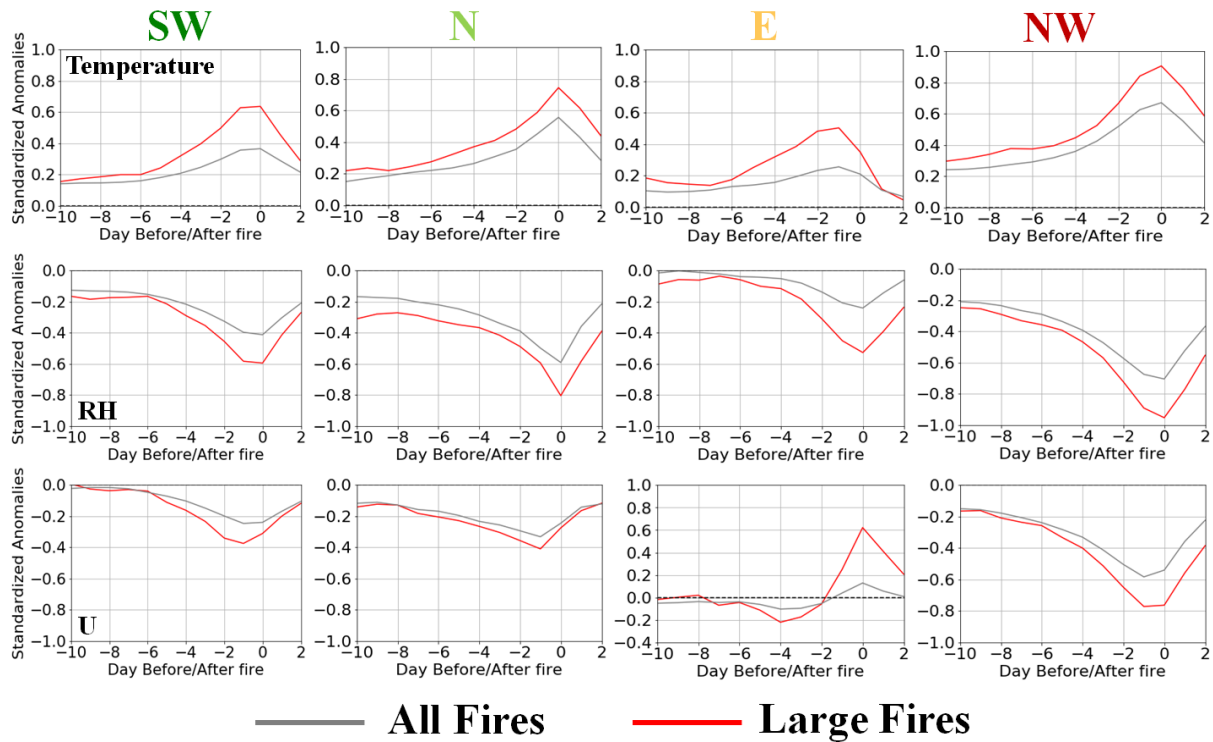


Figure 4.8 - Composites of standardized anomalies of temperature, relative humidity and zonal wind velocity at 12 days for all fires (gray) and for large fires (red) for regions under study.

(12 days), the difference in the distribution of the value of the anomalies for both variables for all fires and LF starts approximately six days before the fire day. The same temporal scale (six days) of significant separation between all fires and LF occurs in the E region, although it has lower anomalies than in the SW region. For the NW region, there is a much-hampered difference between the anomalies for all the fires and the LF than in the SW region. Subsequently, there is a significant increase in the value of the anomalies of temperature and RH, in the case of large fires, with only two days in advance of fire day. Finally, in the N region, although temperature and RH anomalies are always higher in the case of LF, the difference between the anomalies for both classes is constant.

For the indices representing the dryness of the fuels at different layers of the soil (Figure 4.9), it is observed that although the anomalies associated with the LF are always higher (except at SW region), the DMC as the DC does not have significant variations on a scale of 12 days. This behaviour was already expected since the DMC acts to evaluate a dryness of the fuels at 7 cm of depth and the DC at

18 cm of depth, and these indices do not, therefore, reflect the meteorological variations of the day, presenting a temporal lag.

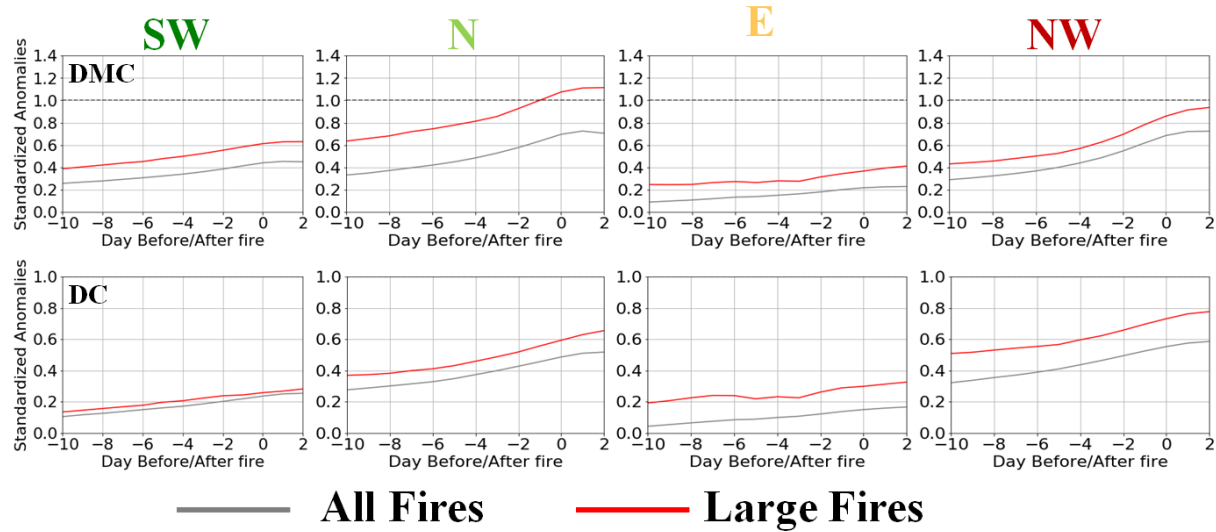


Figure 4.9 - As in Fig. 4.8 but for DMC and DC.

4.3.3. Five Weeks

On a five-weeks scale (four weeks before fire and one week later) the standardized weekly temperature, relative humidity and zonal wind velocity anomalies show diverse anomalies for the different regions (Figure 4.10). The temperature has higher weekly anomalies for the fire week with anomalies ranging from 0.2 to 0.6 standard deviation (std) for LF in the four regions. In the SW, N and E regions, in both temperature and relative humidity, there is an increase in the weekly LF anomalies than those associated with all fires one week before the week of the event. For the NW region, although the anomalies associated with the LF are always higher, there is no clear separation of the distribution of the weekly anomalies (temperature and relative humidity) of the LF and all the fires. Regarding the wind speed, there are no significant variations of the zonal component of the wind on this scale, except for the NW region, where the highest anomalies occur for both all fires and LF (LF reaches a maximum wind anomaly of 0.6 std).

Considering the previous analysis of the 12-day scale for the meteorological variables (section 4.3.2), the non-significance of the results at a 5-weeks scale was already expected, since the influence of the wind is limited to the day of the fire itself. Regarding temperature (and relative humidity) the anomalies are also in line with the results of the previous section with the weekly anomalies for the significant LF for the previous week or for the fire week itself.

The standardized weekly anomalies of the DMC and the DC (Figure 4.11) show that for the four regions considered the LF are associated with positive and higher anomalies of these two indices than when considering all fires. There is, however, an exception for the SW region when considering DC, in which case both fire classes have equal anomalies following the same behaviour of 12 days timescale.

However, this five-weeks time scale does not seem to be the most appropriate as a distinctive factor for the dryness of fuels between the two classes considered. Although the anomalies associated

with the LF are higher, in particular in the N region for the DMC, the slopes of the anomaly curves are the same for the two classes, and it is not possible to determine for this time scale from which moment the anomalies associated with the LF distanced themselves from the other fires.

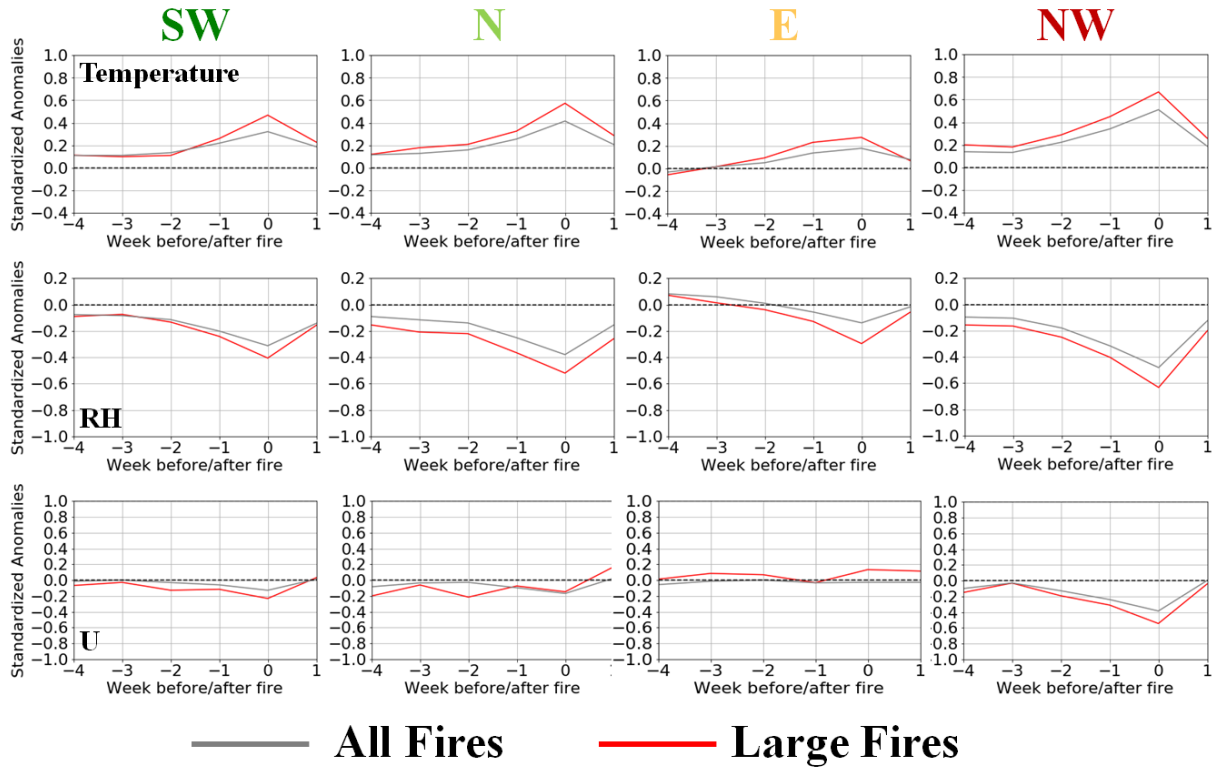


Figure 4.10 - Composites of standardized anomalies of temperature, relative humidity and zonal wind velocity at 5 weeks for all fires (gray) and for large fires (red) for the regions.

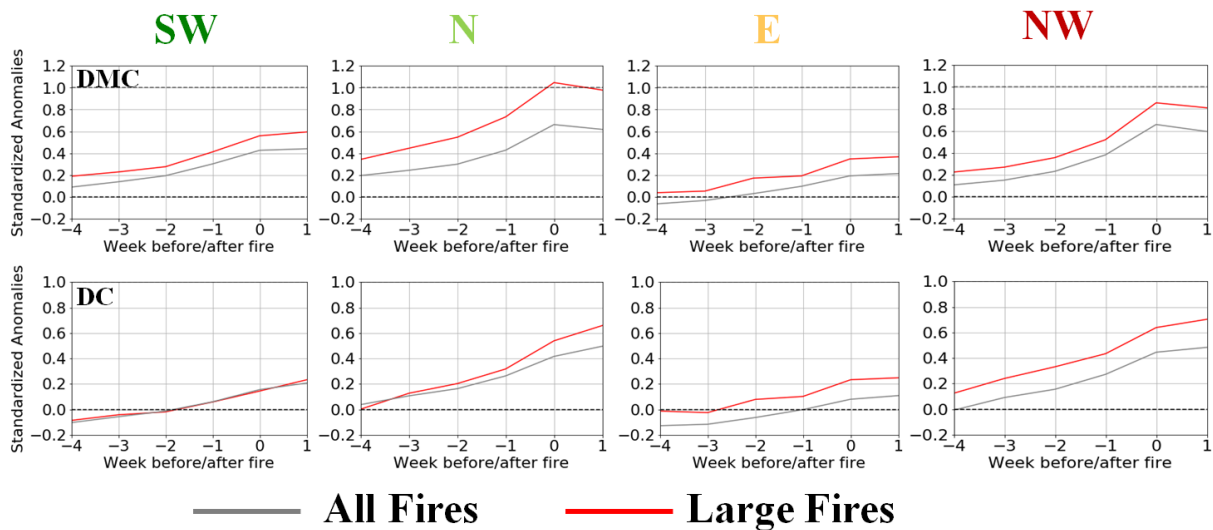


Figure 4.11 - As in Fig. 4.10 but for DMC and DC.

4.3.4. Eight months

At a timescale of 8 months, the standardized monthly anomalies of the meteorological variables are shown in Figure 4.12. For the case of the temperature, it is possible to verify that the monthly anomalies are always below 0.4 std for all the regions and for the two classes in the analysis (all the fires and the LF), where it is emphasized that the highest values occur during the fire season. The anomalies range from negative to positive between one and three months before the fire season considered, depending on the region. For relative humidity, the behaviour is analogous, but with negative anomalies, as previously described in this study. The anomalies on the monthly zonal wind velocity scale remain close to zero for both classes, with a slight peak during the fire season. When analysing the eight months composites, the anomalies decrease immediately after the fire season similarly to what was seen in the other temporal scales. Contrariwise to what happens in the remaining timescales, on the eight months timescale it is difficult to differentiate between the behaviour of LF and all the fires that occur for each one of these four regions.

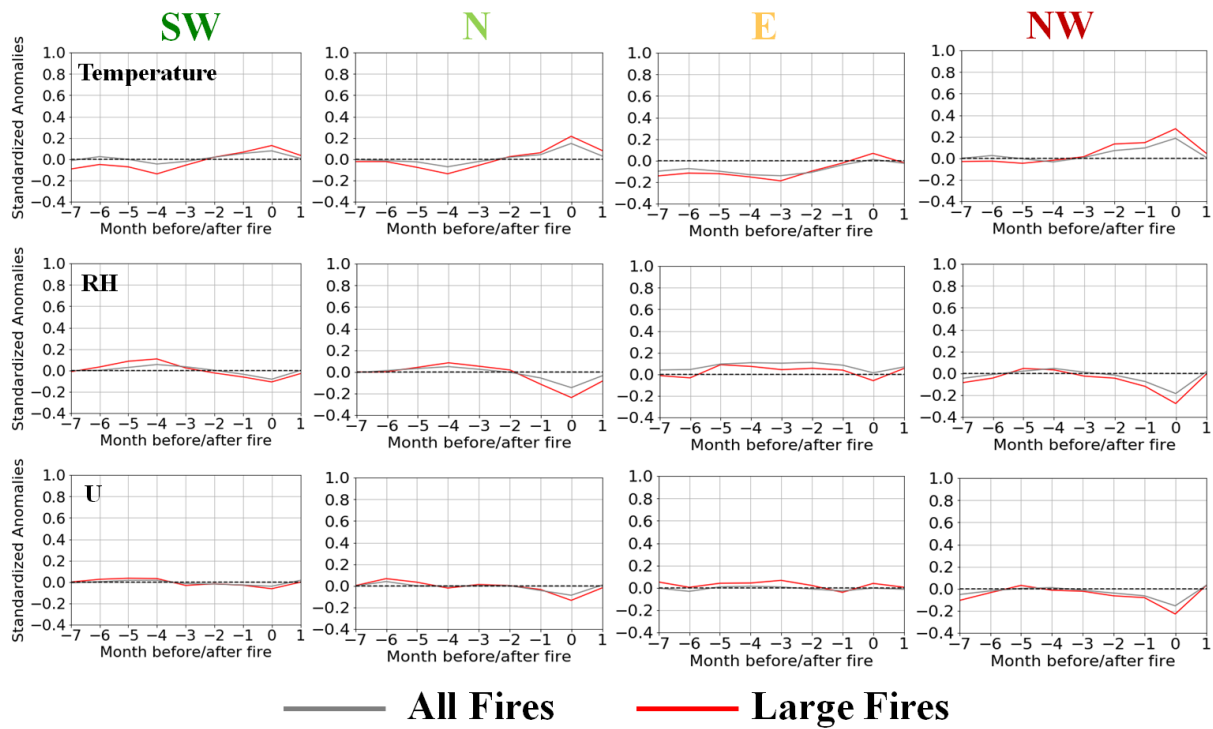


Figure 4.12 - Composites of the monthly standardized anomalies for eight months temperature, relative humidity and zonal wind velocity for all fires (gray) and for large fires (red).

Similarly to the previous timescales, the DMC and the DC composites were calculated for a time scale of 8 months (Figure 4.13). This longer time scale is appropriate to infer the dryness, not only of the fuels at different depths but also of the soil dryness itself, since both variables have as input the precipitation and the temperature, functioning as proxies of the drought and indicative of the pre-season fire conditions.

DMC and DC present rather different behaviour for this time scale. Although with different levels of standardized anomalies, two to three months before the fire season, the DMC anomalies for the LF generally start to rise, a month before the same escalation starts to become evident for the all fires curve.

Moreover, for the N and E regions, there is a clear difference in the distribution between all the fires and the LF from two months before the fire season. Finally, for the NW region, monthly DMC anomalies are practically identical.

The monthly distributions of the DC anomalies of all fires and the LF are not coincident, except for the SW region, and these become positive roughly two months before the fire season. In the N and SW regions, until two months before the fires, the distribution of the monthly anomalies associated with the LF is always below the anomalies where all the fires are included, which could be associated to the fact that the fire season being less dry. Only afterwards there is a reversal of this pattern and, in the previous month to the fire event, the registered DC anomalies for LF are considerably higher than for all fires.

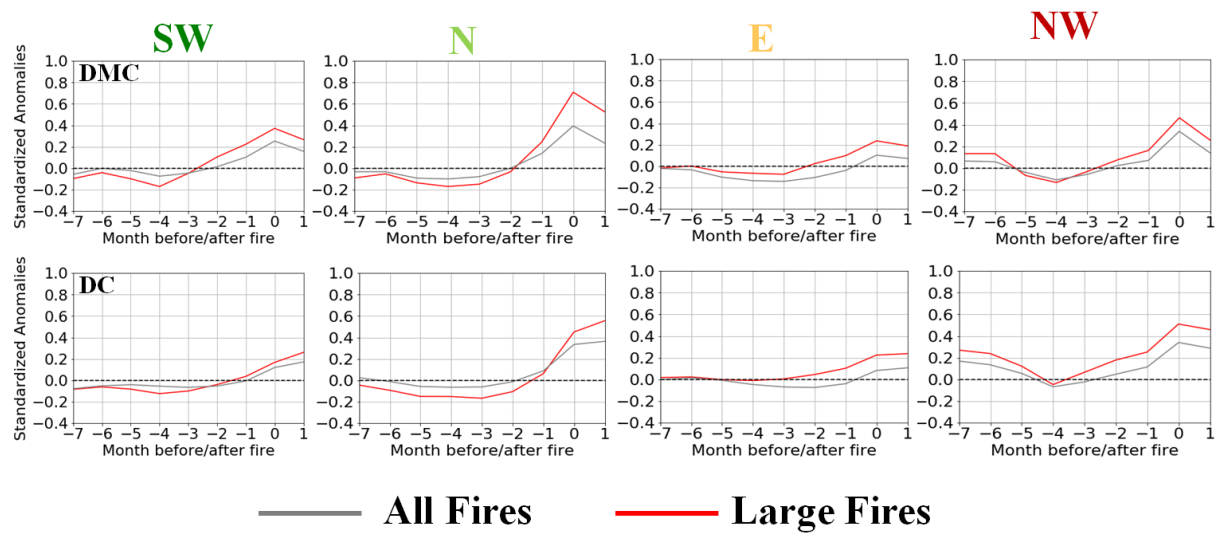


Figure 4.13 - As in Fig. 4.12 but for DMC and DC.

4.4. Principal Component Analysis

. In order to quantify the importance of each of the meteorological variables and the indices that are representative of the dryness of the fuels to the LF activity on the ignition day, a Principal Component Analysis (PCA) was applied to each of the four regions of Iberia.

The explained variance attributed to each principal component is represented in figure 4.14, assigning a different colour in accordance to the meteorological variables depicted: temperature (red), relative humidity (green), zonal wind (blue), DMC (black), DC (pink) and precipitation (yellow), respectively, for the Southwest (SW), North (N), East (E) and Northwest (NW) regions.

For all regions, the variable with higher variance explained for the day of the fire, and thus the variable which dominates the activity associated with the LF is temperature. However, the amount of explained variability expressed by the temperature for each region is different. Temperature is followed by relative humidity (RH), wind direction, DMC, DC and precipitation in all the regions.

Taking into account these results, and in order to reduce the dimensionality of the problem before applying the clusters analysis to each of the regions, it was decided to retain the first three principal components. These components represent the temperature, relative humidity and zonal wind speed, and explain more than 80% of the explained variance of all regions and thus, should be sufficient to identify the various groups of meteorological conditions (FWT) for the LF days. Therefore, the standardized anomalies associated with the days of ignition of temperature, RH and zonal wind velocity were used for each region before applying the *K-Means* algorithm.

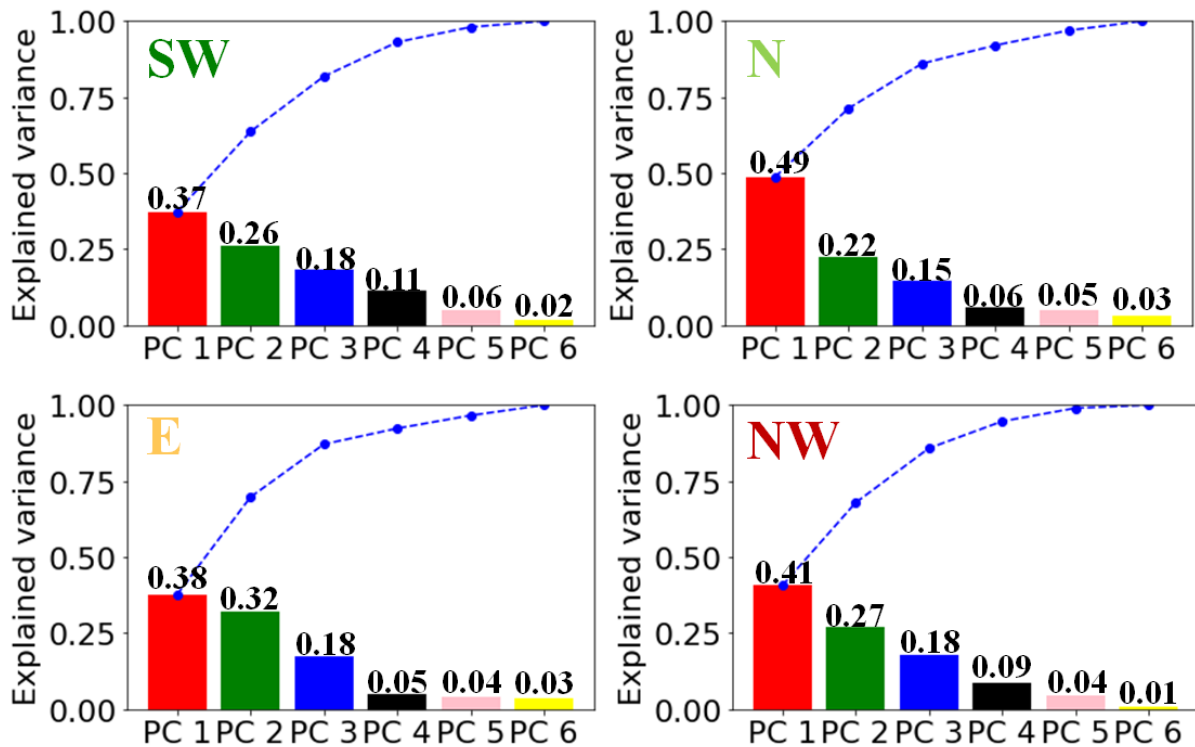


Figure 4.14 – Explained variance for the six principal components (PC 1 through PC 6) corresponding to the temperature (red), relative humidity (green), wind direction (blue), DMC (black), DC (pink) and precipitation (yellow) for the four regions of Iberia. The dashed blue line symbolizes the cumulative explained variance.

4.5. K – Means Analysis: Fire Weather Types identification

Before applying the K-means algorithm, it was necessary to estimate the optimal number of clusters for each of the regions. To achieve this, the gap statistic method was applied, and three clusters were obtained for all regions.

The application of the cluster analysis to aggregate fire day standardized anomalies for the variables identified by the PCA can be observed in figure 4.15, showing that in all four regions, three distinct

FWT were identified. Each FWT is characterized by the final centroid of the standardized anomalies of temperature, relative humidity and zonal wind velocity associated with the LF days.

The FWT_1, depicted in blue, is characterised by high wind speed anomalies (above one std), whereas the FWT_2, in green, is represented by anomalies of the various variables within the normal range (below one std). Finally, FWT_3, in red, is characterized by high positive temperature anomalies (above one std) and strong negative RH anomalies (above one std). However, the value of the anomalies that distinguish each FWT is different for the four regions.

The FWT_1 (blue) for the SW, E and NW regions are characterized by negative standardized anomalies of the zonal wind velocity (close to -1 std) for the first two regions and -1.5 for the later. As in the case of the analysis of the meteorological variables in section 4.2, the negative signal of the zonal wind speed anomalies is indicative of easterly winds. On the other hand, for the Eastern region, FWT_1 presents U anomalies of approximately 2.5 standard deviations, with prevailing winds of the West quadrant. Besides, this FWT is characterized by negative standardized anomalies of relative humidity, in particular, the NW region.

The FWT_2 (green) is characterized by near normal values, thus not configuring any extreme weather conditions since none of the meteorological variables (temperature, relative humidity and zonal wind speed) presents values of anomalies higher than a standard deviation (in absolute value).

Finally, for the FWT_3 (red), all the regions present temperature anomalies for the day of fire close to 1.5 std. In addition to the high-temperature anomalies, and as previously mentioned, the FWT_3 is also distinguishable by strong negative relative humidity anomalies, with all regions presenting values between -1 and -1.5 std.

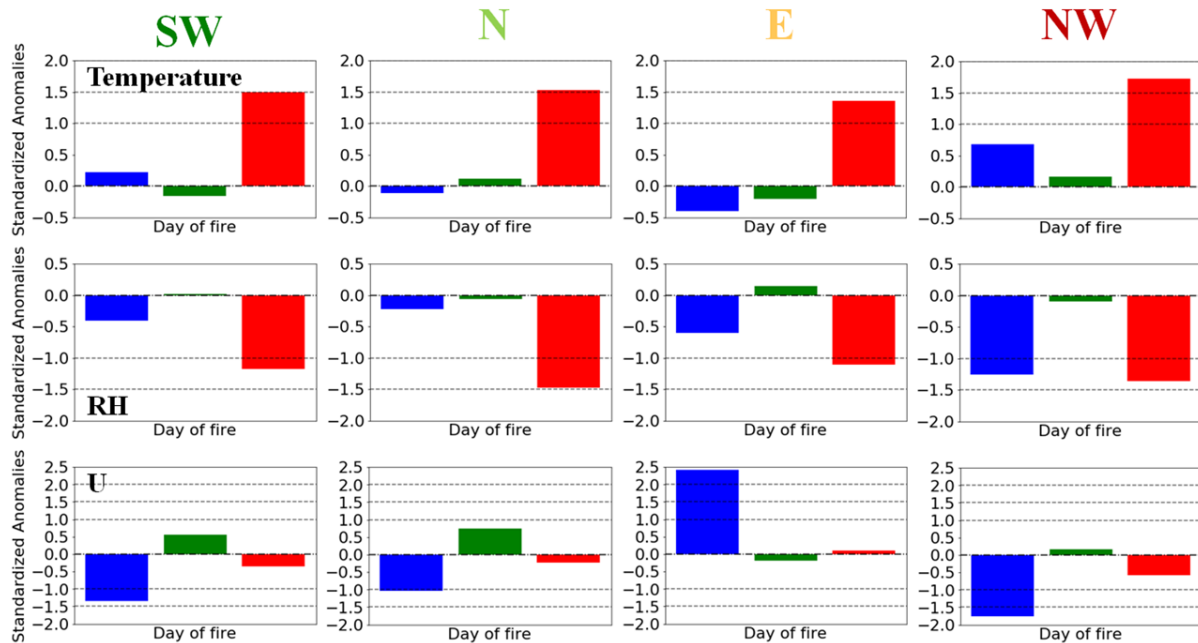


Figure 4.15 - Composites of the standardized anomalies of the meteorological variables (temperature, relative humidity and zonal wind) of the fire day associated with the three Fire Weather Types identified by K-means, FWT_1 (blue), FWT_2 (green) and FWT_3 (red).

The FWTs were characterized, taking into account the meteorological conditions associated with the LF days. However, fuel dryness is an important condition. Therefore, the same procedure was also applied to the fuel dryness associated with each FWT (Figure 4.16). For FWT_1 (blue), both

standardized DMC and DC anomalies are always positive and approximately 0.5 std, except for the NW region, for DC, where the standardized anomalies for the fire day are clearly above 0.5. The FWT_2 (green), for the case, of the DMC always shows positive values, except for the East region that presents slightly negative values. As for the DC case, it presents a behaviour similar to that presented by FWT_1, where NW region is the one with the highest values. Finally, in the case of FWT_3 identified as having above-normal temperature and strong negative relative humidity anomalies, this FWT is the one with the highest standardized anomaly values of DC and DMC for all regions, when compared to the two other FWTs identified. The DMC shows anomalies slightly below (above) 1 std for the SW and E (NW) regions. The Northern region presents the highest values with anomalies of 1.5 std. The DC presents anomalies above 0.5 for all the regions under study, except for the SW region, which presents DC anomalies below 0.5.

As a common denominator, the northernmost regions of IP (NW and N) have higher DMC and DC anomalies than in the other regions, in particular for FWT_3. As it turns out, the monthly average values of these two variables (Figure 4.2) are lower than those of the remaining two regions, which translates into lower average annual temperature and higher precipitation values (Figure 4.1). These two regions are therefore more sensitive to annual variations in precipitation and temperature values than the traditionally driest regions of IP, showing larger variability of DMC and DC values, which translates into highest anomalies associated to LF conditions.

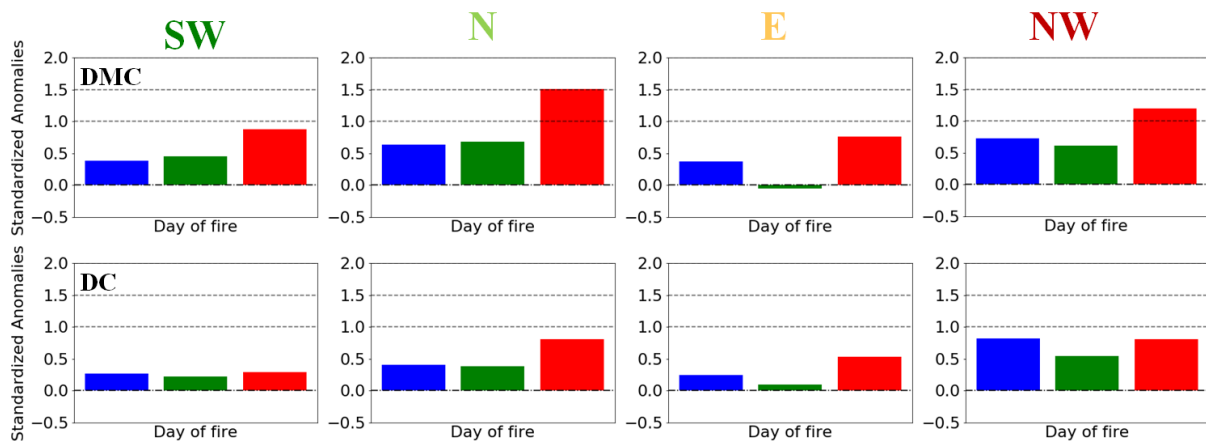


Figure 4.16 - As in Fig. 4.15 but for DMC and DC.

In order to assess the relative importance of the FWTs identified in each region, table 4.2 presents the percentage of events associated with each FWT. This table includes all LF that occurred in the period 1980 to 2015, and the percentage of the LF BA associated to each of the FWTs, taking into account the total burned area for the considered period.

For all regions, FWT_3 is the most frequent pattern associated with the occurrence of LF (Table 4.2). In the N region, it accounts for almost half of all LF, followed by the SW region with about 42% and the E and NW region with 38% and 36%, respectively. However, the FWT associated with high temperature and low relative humidity anomalies is not responsible for the highest amount of burned area in all regions. Thus, in the NW region, the FWT associated to the zonal wind anomalies corresponds to 39% of the total burned area. Finally, the second most frequent FWT varies in the four regions.

Table 4.2 - Representation of the relative importance of the Fire Weather Types identified in each region in terms of the percentage of Large Fires (LF) of the contribution of the respective burned area.

Region	FWT	% LF Events	% Burned Area
SW	FWT_1	25.43	23.45
	FWT_2	32.17	22.60
	FWT_3	42.40	53.95
N	FWT_1	30.03	27.10
	FWT_2	20.80	22.09
	FWT_3	49.17	50.81
E	FWT_1	25.38	20.93
	FWT_2	36.06	23.66
	FWT_3	38.56	55.41
NW	FWT_1	34.08	39.40
	FWT_2	29.44	22.76
	FWT_3	36.48	37.84

4.6. Fire Weather Types characterization

4.6.1. Twelve days

Following the methodology applied in section 4.3, composites of the meteorological variables (temperature, relative humidity and zonal wind velocity) are analyzed again at a 12-days time scale, but in this case, for the three identified FWT (Figure 4.17).

The FWT_1 (blue) as it was already verified in the analysis for the fire day stands out by the high anomalies of zonal wind speed (in absolute value) for all the regions. It is verified that the peak of U is well marked for the fire day (Fig. 4.17), while the days before the fire day present near-zero anomalies and in the two days immediately after the day of the beginning of the fire the anomalies return to the average values. The FWT_2 (green) does not show any pattern in the behaviour of the anomalies for any of the variables under study, and for the 12 days under analysis, the anomalies of these variables are close to zero. The FWT_3 (red) is rather different relatively to the other two FWTs under analysis, with both the temperature and relative humidity showing clear strong anomalies. However, contrary to the wind, the effect of these two variables is not restricted to the fire day but often starts several days before, albeit its influence varies among the four regions. In the SW and E regions for both temperature and RH, the distribution of the anomaly values for FWT_3 is separated from the other two FWT up to six days before the fire day where the maximum value of the anomaly is reached and decreasing on the two days immediately after the fire day. For the N region, the distribution of the anomaly values for FWT_3 differentiates from the other two FWT, since four days before the fire day. At a 12-day time scale, both

anomalies associated with FWT of the indices representing fuel dryness are positive for all regions, except for the eastern region for FWT_2 (Figure 4.18). FWT_3 presents a significant higher DMC and DC anomalies for all regions (except the DC for the NW region). The difference between the FWT remains practically the same because the values of DMC and DC do not change significantly at this timescale, as explained in section 4.2.

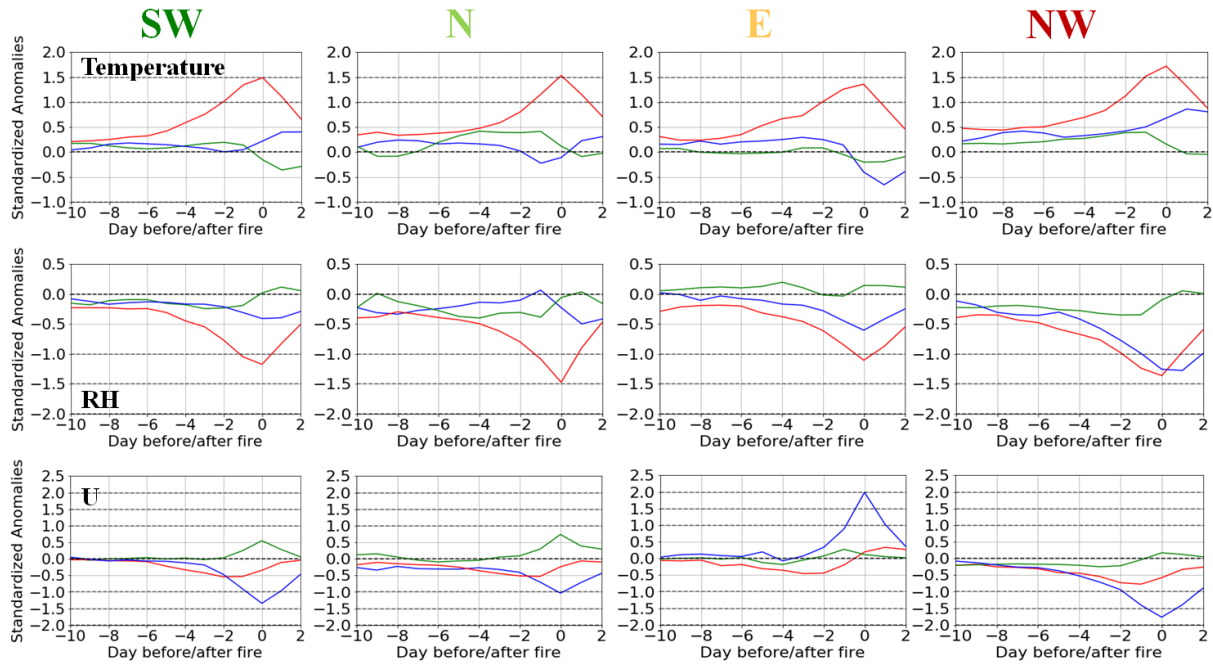


Figure 4.17 - Composites of the standardized anomalies of the meteorological variables (temperature, relative humidity and zonal wind) at 12 days timescale associated with the three Fire Weather Types identified by K-means, FWT_1 (blue), FWT_2 (green) and FWT_3 (red).

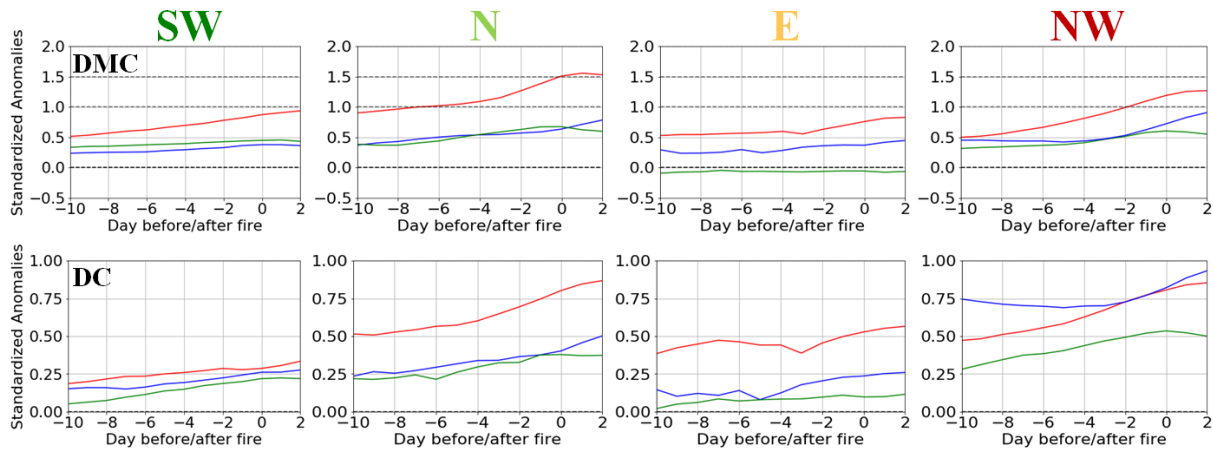


Figure 4.18 – As in Fig. 4.17 but for DMC and DC.

4.6.2. Eight months

At an eight-months scale, the monthly standardized anomalies of the meteorological variables for the LF associated with each of the three FWTs identified are only significant for the month in which the fire occurred, as it has already been pointed out before.

Therefore, only the monthly standardized anomalies of the DMC and DC of the three FWTs were represented (Figure 4.19).

For the DMC of the SW region it is possible to observe that up to four months before the fire season all FWT have slightly negative anomalies, however, four months before the FWT_3, it begins to stand out from the rest with higher positive anomalies reaching its peak during the fire season and descending in the following month. Regarding DC, the behaviour is different, with the FWT_3 being the one with the most negative DC anomalies up to four months before the fire season. However, one month before the DC anomalies become positive, and during the fire season, this presents the highest values compared to FWT_1 and FWT_2. Interestingly, on the contrary to the DMC for this region, the DC anomaly values continue to increase in the month after the fire. This behaviour is partly due to the lag time of DC being higher than that of DMC but also to the fact that not enough precipitation occurs to attenuate fuel dryness right after the month of fire.

In the N (E) region, the DMC presents slightly negative anomalies for all FWT up to 2 months before the fire season. In the two months preceding the fire month, the monthly DMC anomalies for FWT_3 start to grow much faster than that of the remaining FWT, reaching in the fire month an anomaly slightly higher than one (0.5) standard deviation and decreasing immediately in the month following the month of fire. The DC exhibits the same behaviour as the DMC, where two months before the fire season, DC anomalies become positive for all FWTs, but FWT_2 has higher values. As in the SW region, DC values continue to rise in the month following the month of fire or remain the same for the E region.

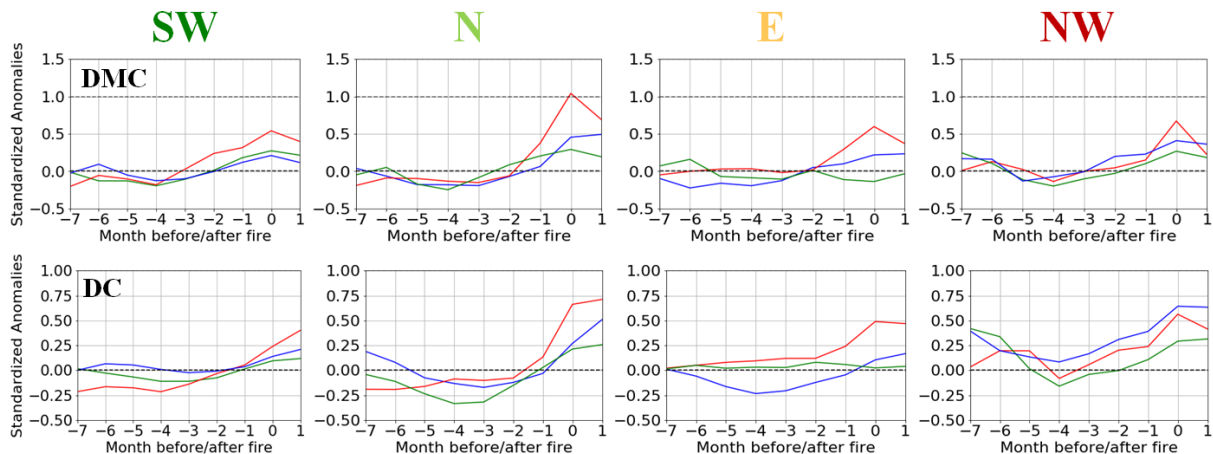


Figure 4.19 - Composites of the standardized anomalies of DC and DMC at 8 months timescale associated with the three FWT.

For the DMC in the NW region, it is not possible to distinguish between different FWT. It is verified that up to three (two) months before the fire season the DMC anomalies associated with FWT_1 and FWT_3 (FWT_2) are close to zero and that afterwards the anomalies become positive and grow until they reach its maximum during the fire season, particularly for FWT_3. Unlike the other regions of the IP, the monthly anomalies associated with the LF of FWT_1 present the highest values of DC,

although, in the case of FWT_3, the anomalies have values in the order of 0.1 lower when compared with FWT_1. For all FWTs the DC anomalies begin to grow more rapidly four months before the month of fire.

5. Conclusions

This study proposes to analyse historical meteorological data and fire records aiming to classify large summer fires for four regions of Iberia according to their local-scale weather conditions (i. e. temperature, relative humidity, wind speed) and fire danger weather indices that are components of the Canadian FWI (duff moisture code and drought code).

Based on the historical data, LF were identified based on a threshold approach for four different regions of the Iberian Peninsula. The different threshold values for each region support the hypothesis that each region is characterized by a specific fire regime which causes that LF to be triggered by different conditions.

In order to identify the different climatic conditions associated with the occurrence of LF, it is necessary to understand the importance of the different variables which are related to the occurrence of the LF at different time scales. At a 12-day scale, the meteorological variables are those that represent greater importance in distinguishing the activity associated with LF, but with different lags for the regions under study. The zonal wind has a restricted influence on the fire day. The anomalies of temperature and RH has a significant separation between all fires and LF, although with different lags for the four regions. For eight months, that is, on a scale that is compatible to study the of interannual variations of the variables and the conditions pre-fire season, the indices that translate the dryness of the fuels to different layers of the soil (proxies of drought conditions) are those with a significant difference between the activity associated to all fires and LF, although at different lags for the different regions.

The fires considered for the extended fire season (May to October) generally occur when two conditions come together:

- 1) very dry fuels are present, or in a drought situation, processes that as we have seen act on a monthly scale;
- 2) when certain meteorological conditions are met (acting at the synoptic scale).

Based on the results obtained in the composite analysis, a PCA was applied to standardized anomalies associated with LF days of meteorological variables and fuel dryness indices allowing to reduce the dimensionality of the problem to only three variables (temperature, relative humidity and zonal wind) that explain at least 80% of the variance associated with LF days. Based on these three variables, three FWT were identified based on the synoptic conditions associated with LF days, each of those having distinct meteorological anomalies at different timescales.

The first FWT identified, designated in this study as FWT_1, is mainly characterized by significant zonal wind speed anomalies for all regions and moderate (below a standard deviation) RH anomalies, except for the NW region and is also characterized by significant HR anomalies. This FWT is associated to the LF days presenting risks due to the wind speed and intensity (Ruffault et al., 2016, 2018). In the Iberian Peninsula, the LF days associated with this FWT present a two distinguish

behaviours. Hoinka et al. (2009) evaluated the evolution of the synoptic and mesoscale wind, temperature and humidity pattern during wildland fire events above 500 ha in Portugal. This study identifies that for the fire day, most of these fires present winds with east component transported from inside Spain. Rasilla et al. (2010) showed that the E region of the IP is characterized by westerly winds which contribute to the displacement of air masses from the Atlantic which crosses over the Spanish mainland, suffering a strong drying and warming that increases the risk of wildfire. These two studies confirm the two predominant wind regimes identified in the IP associated with the days of LF. The SW, N and NW region of the Iberian Peninsula are essentially associated with easterly winds while the E region is mostly affected by westerly winds. Contrary to what was observed in the Mediterranean region of France (Ruffault et al. 2016), this FWT is not predominant in the IP, however, for the NW region, the FWT_1 is responsible for a larger amount of BA associated with the LF events.

The FWT_2 presents anomalies of the meteorological variables within the normal range, thus not presenting any major specific atmospheric characteristic, although it is characterized by positive anomalies of the indices representative of the dryness of the fuels for the whole IP. The fact that LF occurs without an exceptional condition demonstrates the difficulty of the connection between fire occurrence, weather and climate. There is a multitude of other previously identified factors besides weather that influence the development of fires. This FWT_2 accounting for almost a quarter of LF BA in IP needs further study in the future.

The meteorological characteristics associated with FWT_3, namely high temperature and low relative humidity, are described as one of the predominant conditions associated with the occurrence of LF in IP. Pereira et al. (2005) have shown that LF in Portugal is related to anomalous advection of hot and dry air from Northern Africa and through central Iberia, associated with high temperatures and low RH. Also, in Trigo et al. (2006) the authors have shown that the exceptional fire season of 2003 in Portugal was characterized by significant anomalies of surface meteorological variables, namely surface maximum and minimum temperature and relative humidity. In fact, and unlike other Mediterranean regions, such as Southern France (Ruffault et al. 2016, 2018), in the IP this FWT is the one that most contributes to the occurrence of LF. Besides the meteorological variables, FWT_3 is the one with the highest anomalies of DMC and DC, especially for the NW region. This region of IP is associated with colder and wetter climate, a fact visible on the behaviour of DC where one month after the month of fire (eight months timescale) the anomalies drop. This region perhaps is more susceptible to the annual variability of these variables than Eastern regions traditionally drier and were after the fire season the anomalies of DC continues to rise, illustrates the temporal lag of the variable but in particular the very dry conditions of this region.

The projected rise in temperature and droughts frequency, as well as a decrease in precipitation for the Mediterranean region (IPCC SR1.5, 2018), could lead in the future to an increase of heatwaves, which are characterized by days in with above-normal temperatures. Since FWT_3 is the most important FWT in IP associated with the development of LF this could lead to an increase of LF events. However, trends in the number of annual LFs associated with FWT_3 are either not significant for the SW and NW regions or have a negative trend in the remaining regions (Figure A.3). In fact, for regions that include only the Spanish territory (N and E regions), the trend is negative, which probably reflects the effort of human fire suppression policies that have been applied in Spanish territory after the multiple LF of the 1980s (Viegas D. X., 2014). For regions which also includes Portuguese LF trends for FWT_3 has no significant changes.

The approach followed in this study is partially based on the methodology proposed by Ruffault et al. (2016) which allowed to objectively identify for the different regions of Iberia multiple fire climatologies associated with the occurrence of LF. Nevertheless, we would like to highlight that the area analysed is far greater and affected by different conditions comparatively to the one studied by

Ruffault et al. (2016). Following the methodology described here, it is possible to identify the climatic conditions that distinguish the occurrence of the LF events at different time scales, permitting to identify for the four regions different time lags for the variables under study.

As future work, the drivers associated with FWT_2 may be studied in greater detail, since, despite drought conditions, meteorological factors do not seem to be sufficient to explain the LF classified in this FWT. Also, these different identified climatologies may be included in future projections in order to gauge the variability of these variables to improve knowledge about future conditions associated with LF at the Iberian Peninsula level.

6. References

- Amraoui, M., Pereira, M. G., DaCamara, C. C., & Calado, T. J. (2015). “Atmospheric conditions associated with extreme fire activity in the Western Mediterranean region”. *Science of the Total Environment* 524–525, 32–39. <https://doi.org/10.1016/j.scitotenv.2015.04.032>.
- Anil K. Jain. (2010). “Data clustering: 50 years beyond K-means”. *Pattern Recognition Letters* 31(8), 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>.
- Archibald, S., Lehmann, C. E. R., Gomez-Dans, J. L., & Bradstock, R. A. (2013). “Defining pyromes and global syndromes of fire regimes”. *Proceedings of the National Academy of Sciences* 110(16), 6442–6447. <https://doi.org/10.1073/pnas.1211466110>.
- Archibald, S., Roy, D. P., van Wilgen, B. W., & Scholes, R. J. (2009). “What limits fire? An examination of drivers of burnt area in Southern Africa”. *Global Change Biology* 15(3), 613–630. <https://doi.org/10.1111/j.1365-2486.2008.01754.x>.
- Barbero, R., Abatzoglou, J. T., Kolden, C. A., Hegewisch, K. C., Larkin, N. K., & Podschwit, H. (2015). “Multi-scalar influence of weather and climate on very large-fires in the Eastern United States”. *International Journal of Climatology* 35(8), 2180–2186. <https://doi.org/10.1002/joc.4090>.
- Barbero, R., Abatzoglou, J. T., Larkin, N. K., Kolden, C. A., & Stocks, B. (2014). “Climate change presents increased potential for very large fires in the contiguous United States”. *International Journal of Wildland Fire* 24(7), 892–899. <https://doi.org/10.1071/WF15083>.
- Berrisford, P., Kallberg, P., Kobayashi, S., Dee, D., Uppala, S., Simmons, A. J., & Sato, H. (2011). “The ERA-Interim archive version 2.0.” *ERA Report Series*. ECMWF.
- Bowman, D.M.J.S., Balch, J.K., Artaxo, P., Bond, W.J., Carlson, J.M., Cochrane, M.A., D’Antonio, C.M., DeFries, R.S., Doyle, J.C., Harrison, S.P., et al. (2009). “Fire in the earth system.” *Science* 324, 481–484. <https://doi.org/10.1126/science.1163886>.
- Bradstock, R. A., Cohn, J. S., Gill, A. M., Bedward, M., & Lucas, C. (2009). “Prediction of the probability of large fires in the Sydney region of south-eastern Australia using fire weather”. *International Journal of Wildland Fire* 18(8), 932–943. <https://doi.org/10.1071/WF08133>.
- Ciuti, S., Tripke, H., Antkowiak, P., Gonzalez, R. S., Dormann, C. F., & Heurich, M. (2018). “An efficient method to exploit LiDAR data in animal ecology. *Methods in Ecology and Evolution*” 9(4), 893–904. <https://doi.org/10.1111/2041-210X.12921>.
- Copernicus Emergency Management Service (2019). Retrieved from <http://effis.jrc.ec.europa.eu/>.
- Costa, L., Thonicke, K., Poulter, B., & Badeck, F. (2011). “Sensitivity of Portuguese forest fires to climatic, human, and landscape variables: Subnational differences between fire drivers in extreme fire years and decadal averages”. *Regional Environmental Change* 11(3), 543–551. <https://doi.org/10.1007/s10113-010-0169-6>.
- De La Riva, J., Pérez-Cabello, F., Lana-Renault, N., & Koutsias, N. (2004). “Mapping wildfire occurrence at regional scale”. *Remote Sensing of Environment* 92(2), 288–294. <https://doi.org/10.1016/j.rse.2004.06.013>.
- Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, et al. (2011). “The ERA-Interim reanalysis: Configuration and performance of the data assimilation system”. *Quarterly Journal of the Royal Meteorological Society* 137(656), 553–597. <https://doi.org/10.1002/qj.828>.

- Erickson, M. J., Charney, J. J., & Colle, B. A. (2016). "Development of a fire weather index using meteorological observations within the northeast United States". *Journal of Applied Meteorology and Climatology* 55(2), 389–402. <https://doi.org/10.1175/JAMC-D-15-0046.1>.
- Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Édouard Duchesnay. (2011). "Scikit-learn: Machine Learning in Python" 12, 2825–2830.
- Field, R. D., A. C. Spessa, N. A. Aziz, A. Camia, A. Cantin, R. Carr, W. J. De Groot, et al. (2015). "Development of a Global Fire Weather Database". *Natural Hazards and Earth System Sciences* 15(6), 1407–1423. <https://doi.org/10.5194/nhess-15-1407-2015>.
- Flannigan, M. D., Stocks, B. J., & Wotton, B. M. (2000). "Climate change and forest fires". *Science of the Total Environment* 262(3), 221–229. [https://doi.org/10.1016/S0048-9697\(00\)00524-6](https://doi.org/10.1016/S0048-9697(00)00524-6).
- Ganteaume, A., & Jappiot, M. (2013). "What causes large fires in Southern France". *Forest Ecology and Management* 294, 76–85. <https://doi.org/10.1016/j.foreco.2012.06.055>.
- Ganteaume, A., Camia, A., Jappiot, M., San-Miguel-Ayanz, J., Long-Fournel, M., & Lampin, C. (2013). "A review of the main driving factors of forest fire ignition over Europe". *Environmental Management* 51, 651–662 <https://doi.org/10.1007/s00267-012-9961-z>.
- Hoinka, K. P., Carvalho, A., & Miranda, A. I. (2009). "Regional-scale weather patterns and wildland fires in central Portugal". *International Journal of Wildland Fire* 18(1), 36–49. <https://doi.org/10.1071/WF07045>.
- IPCC. (2018). "Global Warming of 1.5 °C". Retrieved from https://www.ipcc.ch/site/assets/uploads/sites/2/2018/07/SR15_SPM_version_stand_alone_LR.pdf <http://www.ipcc.ch/report/sr15>.
- Jin X., Han J. (2011). "K-Means Clustering". In: Sammut C., Webb G.I. (eds) *Encyclopedia of Machine Learning*. Springer, Boston, MA. <https://doi.org/10.1007/978-0-387-30164-8>.
- Jolliffe, I. T. (2002). "Principal Component Analysis", Second Edition. *Encyclopedia of Statistics in Behavioral Science*, 30(3), 487. <https://doi.org/10.2307/1270093>.
- Kasischke, E. S., Williams, D., & Barry, D. (2002). "Analysis of the patterns of large fires in the boreal forest region of Alaska". *International Journal of Wildland Fire* 11(2), 131–144. <https://doi.org/10.1071/WF02023>.
- Kassomenos, P., Vardoulakis, S., Borge, R., Lumbreras, J., Papaloukas, C., & Karakitsios, S. (2010). "Comparison of statistical clustering techniques for the classification of modelled atmospheric trajectories". *Theoretical and Applied Climatology* 102(1), 1–12. <https://doi.org/10.1007/s00704-009-0233-7>.
- Lawrence, M. G. (2005). "The relationship between relative humidity and the dewpoint temperature in moist air: A simple conversion and applications". *Bulletin of the American Meteorological Society* 86(2), 225–233. <https://doi.org/10.1175/BAMS-86-2-225>.
- Malik, M. R., Isaac, B. J., Coussement, A., Smith, P. J., & Parente, A. (2018). "Principal component analysis coupled with nonlinear regression for chemistry reduction". *Combustion and Flame* 187, 30–41. <https://doi.org/10.1016/j.combustflame.2017.08.012>.
- Moreno, M. V., Malamud, B. D., & Chuvieco, E. (2011). "Wildfire frequency-area statistics in Spain". In *Procedia Environmental Sciences* 7, 182–187. <https://doi.org/10.1016/j.proenv.2011.07.032>.
- Navas, M., & Ordonez, C. (2009). "Efficient computation of PCA with SVD in SQL". *Association for Computing Machinery* 1-10. <https://doi.org/10.1145/1581114.1581119>.
- Nikkhah, A., Rohani, A., Rosentrater, K. A., El Haj Assad, M., & Ghnimi, S. (2019). "Integration of principal component analysis and artificial neural networks to more effectively predict

- agricultural energy flows”. *Environmental Progress and Sustainable Energy*. <https://doi.org/10.1002/ep.13130>.
- Olawale, F., & Garwe, D. (2010). “Obstacles to the growth of new SMEs in South Africa : A principal component analysis approach”. *African Journal of Business Management* 4(5), 729–738.
- Ozturk, T., Ceber, Z. P., Türkeş, M., & Kurnaz, M. L. (2015). “Projections of climate change in the Mediterranean Basin by using downscaled global climate model outputs”. *International Journal of Climatology* 35(14), 4276–4292. <https://doi.org/10.1002/joc.4285>.
- Parente, J., Pereira, M. G., & Tonini, M. (2016). “Space-time clustering analysis of wildfires: The influence of dataset characteristics, fire prevention policy decisions, weather and climate”. *Science of the Total Environment* 559, 151–165. <https://doi.org/10.1016/j.scitotenv.2016.03.129>.
- Pausas, J. G., & Vallejo, V. R. (2011). “The role of fire in European Mediterranean ecosystems”. *Remote Sensing of Large Wildfires*, 3–16. https://doi.org/10.1007/978-3-642-60164-4_2.
- Peña-Gallardo, M., Vicente-Serrano, S. M., Hannaford, J., Lorenzo-Lacruz, J., Svoboda, M., Domínguez-Castro, F., Kenawy, A. E. (2019). “Complex influences of meteorological drought time-scales on hydrological droughts in natural basins of the contiguous Unites States”. *Journal of Hydrology*. <https://doi.org/10.1016/j.jhydrol.2018.11.026>.
- Pereira, M. G., Malamud, B. D., Trigo, R. M., & Alves, P. I. (2011). “The history and characteristics of the 1980-2005 Portuguese rural fire database”. *Natural Hazards and Earth System Science* 11(12), 3343–3358. <https://doi.org/10.5194/nhess-11-3343-2011>.
- Pereira, M. G., Trigo, R. M., Da Camara, C. C., Pereira, J. M. C., & Leite, S. M. (2005). “Synoptic patterns associated with large summer forest fires in Portugal”. *Agricultural and Forest Meteorology* 129(1–2), 11–25. <https://doi.org/10.1016/j.agrformet.2004.12.007>.
- Pinto, M. M., DaCamara, C. C., Trigo, I. F., Trigo, R. M., Turkman, K. F. (2018). “Fire danger rating over Mediterranean Europe based on fire radiative power derived from Meteosat”. *Natural Hazards and Earth System Sciences* 18(2), 515–529. <https://doi.org/10.5194/nhess-18-515-2018>.
- Polade, S. D., Gershunov, A., Cayan, D. R., Dettinger, M. D., & Pierce, D. W. (2017). “Precipitation in a warming world: Assessing projected hydro-climate changes in California and other Mediterranean climate regions”. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-11285-y>.
- Rasilla, D. F., García-Codron, J. C., Carracedo, V., & Diego, C. (2010). “Circulation patterns, wildfire risk and wildfire occurrence at continental Spain”. *Physics and Chemistry of the Earth* 35(9–12), 553–560. <https://doi.org/10.1016/j.pce.2009.09.003>.
- Riley, K. L., Abatzoglou, J. T., Grenfell, I. C., Klene, A. E., & Heinsch, F. A. (2013). “The relationship of large fire occurrence with drought and fire danger indices in the western USA, 1984-2008: the role of temporal scale”. *International Journal of Wildland Fire* 22(7), 894–909. <https://doi.org/10.1071/WF12149>.
- Rodrigues, M., & De la Riva, J. (2014). “An insight into machine-learning algorithms to model human-caused wildfire occurrence”. *Environmental Modelling and Software* 57, 192–201. <https://doi.org/10.1016/j.envsoft.2014.03.003>.
- Ruffault, J., Curt, T., Martin-Stpaul, N. K., Moron, V., & Trigo, R. M. (2018). “Extreme wildfire events are linked to global-change-type droughts in the northern Mediterranean”. *Natural Hazards and Earth System Sciences* 18(3), 847–856. <https://doi.org/10.5194/nhess-18-847-2018>.
- Ruffault, J., Moron, V., Trigo, R. M., & Curt, T. (2016). “Objective identification of multiple large fire climatologies: An application to a Mediterranean ecosystem”. *Environmental Research Letters* 11(7). <https://doi.org/10.1088/1748-9326/11/7/075006>.

- Sánchez-Benítez, A., García-Herrera, R., Barriopedro, D., Sousa, P. M., & Trigo, R. M. (2018). “June 2017: The Earliest European Summer Mega-heatwave of Reanalysis Period”. *Geophysical Research Letters* 45(4), 1955–1962. <https://doi.org/10.1002/2018GL077253>.
- Santos, J. A., & Belo-Pereira, M. (2019). “A comprehensive analysis of hail events in Portugal: Climatology and consistency with atmospheric circulation”. *International Journal of Climatology* 39(1), 188–205. <https://doi.org/10.1002/joc.5794>.
- Saraswathi, S., & Sheela, M. I. (2014). “A Comparative Study of Various Clustering Algorithms in Data Mining”. *International Journal of Computer Science and Mobile Computing* 3 (11), 422–428.
- Scikit-learn user guide. (2019). Retrieved from <https://scikit-learn.org/0.20/downloads/scikit-learn-docs.pdf>.
- Skinner, W. R., Flannigan, M. D., Stocks, B. J., Martell, D. L., Wotton, B. M., Todd, J. B., Bosch, E. M. (2002). “A 500 hPa synoptic wildland fire climatology for large Canadian forest fires 1959–1996”. *Theoretical and Applied Climatology* 71(3–4), 157–169. <https://doi.org/10.1007/s007040200002>.
- Sousa, P. M., Trigo, R. M., Pereira, M. G., Bedia, J., & Gutiérrez, J. M. (2015). “Different approaches to model future burnt area in the Iberian Peninsula”. *Agricultural and Forest Meteorology* 202, 11–25. <https://doi.org/10.1016/j.agrformet.2014.11.018>
- Sumathi, S., & Sivanandam, S. N. (2006). “Introduction to Data Mining and its Applications”. Springer Berlin Heidelberg. 29. <https://doi.org/10.1007/978-3-540-34351-6>.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2005). “Introduction to Data Mining”. [https://doi.org/10.1016/0022-4405\(81\)90007-8](https://doi.org/10.1016/0022-4405(81)90007-8).
- Tibshirani, R., Walther, G., & Hastie, T. (2001). “Estimating the number of clusters in a data set via the gap statistic”. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 63(2), 411–423. <https://doi.org/10.1111/1467-9868.00293>.
- Trigo, I. F., Davies, T. D., & Bigg, G. R. (1999). “Objective climatology of cyclones in the Mediterranean region”. *Journal of Climate* 12(6), 1685–1696. [https://doi.org/10.1175/1520-0442\(1999\)012](https://doi.org/10.1175/1520-0442(1999)012).
- Trigo, R. M., Pereira, J. M. C., Pereira, M. G., Mota, B., Calado, T. J., Dacamara, C. C., & Santo, F. E. (2006). “Atmospheric conditions associated with the exceptional fire season of 2003 in Portugal”. *International Journal of Climatology* 26(13), 1741–1757. <https://doi.org/10.1002/joc.1333>.
- Trigo, R. M., Sousa, P. M., Pereira, M. G., Rasilla, D., & Gouveia, C. M. (2013). “Modelling wildfire activity in Iberia with different atmospheric circulation weather types”. *International Journal of Climatology* 36(7), 2761–2778. <https://doi.org/10.1002/joc.3749>.
- Turco, M., Llasat, M. C., von Hardenberg, J., & Provenzale, A. (2013). “Impact of climate variability on summer fires in a Mediterranean environment (northeastern Iberian Peninsula)”. *Climatic Change* 116(3–4), 665–678. <https://doi.org/10.1007/s10584-012-0505-6>.
- Turco, M., Von Hardenberg, J., AghaKouchak, A., Llasat, M. C., Provenzale, A., & Trigo, R. M. (2017). “On the key role of droughts in the dynamics of summer fires in Mediterranean Europe”. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-00116-9>.
- Unglert, K., Radić, V., & Jellinek, A. M. (2016). “Principal component analysis vs. self-organizing maps combined with hierarchical clustering for pattern recognition in volcano seismic spectra”. *Journal of Volcanology and Geothermal Research* 320, 58–74. <https://doi.org/10.1016/j.jvolgeores.2016.04.014>.
- Urbieto, Itziar R., Gonzalo Zavala, Joaquín Bedia, José M. Gutiérrez, Jesús San Miguel-Ayanz, Andrea Camia, Jon E. Keeley, and José M. Moreno. (2015). “Fire activity as a function of fire-weather seasonal severity and antecedent climate across spatial scales in southern Europe and Pacific

- western USA”. *Environmental Research Letters*, 10(11). <https://doi.org/10.1088/1748-9326/10/11/114013>.
- Van Wagner, C. E. (1987). “Development and structure of the Canadian forest fire weather index system. Forestry”. Chalk River, Ontario: Environment Canada, Canadian Forest Service, Petawawa Forest Experiment Station, Forest Technical Report 35. <https://doi.org/19927>.
- Vázquez, A., & Moreno, J. M. (1995). “Patterns of Fire Occurrence Across a Climatic Gradient and Its Relationship to Meteorological Variables in Spain”. 408–434. https://doi.org/10.1007/978-1-4612-4186-7_20.
- Velez R. (2011). “Fire Situation in Spain”. *Global Forest Fire Assessment 1990-2001*. Roma: FAO.
- Viegas, D. X. (2014). “Advances in forest fire research”. *Imprensa da Universidade de Coimbra*. <https://doi.org/10.14195/978-989-26-0884-6>.
- Wilks, D. S. (2011). “Statistical methods in the atmospheric sciences”. *Academic press* 100, 301–394.

7. Annexes

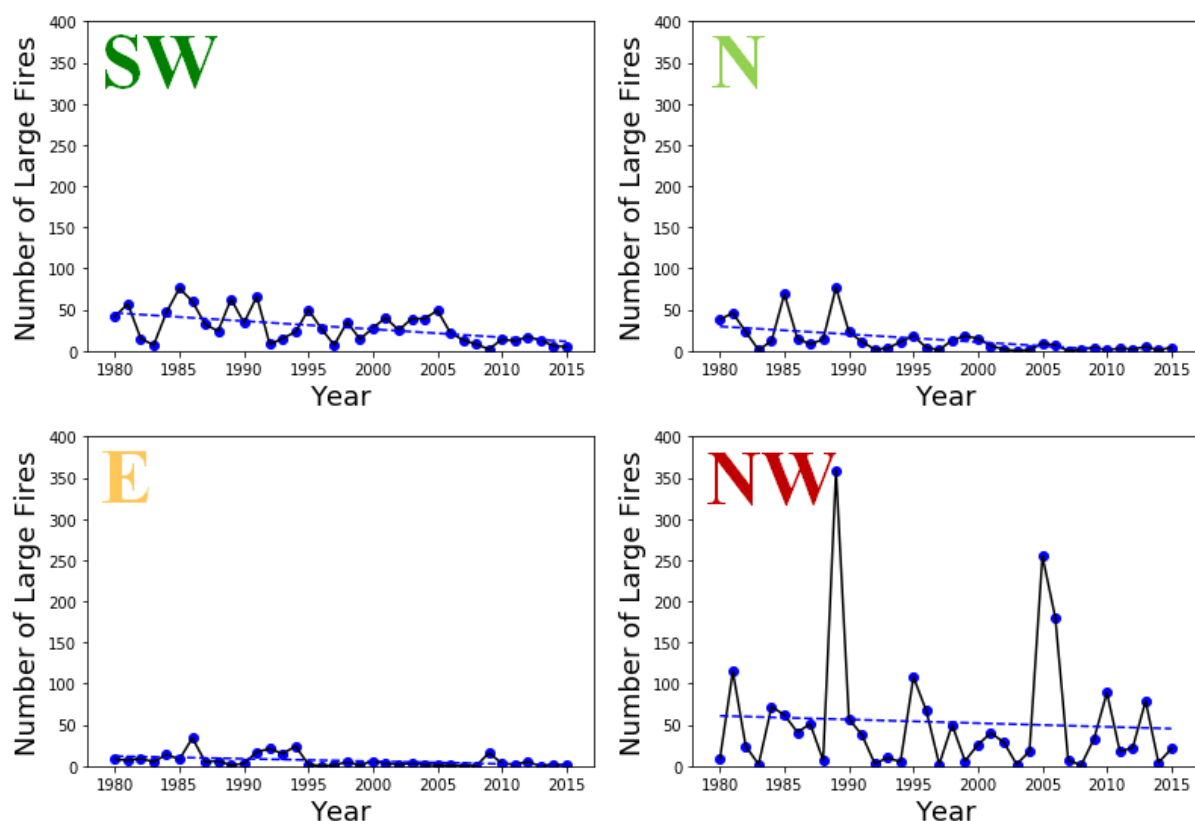


Figura A.1 – Trends of FWT_1 annual number of large fires (Mann- Kendall Test $p < 0.001$) for the SW, N and E with a negative trend and NW with no trend.

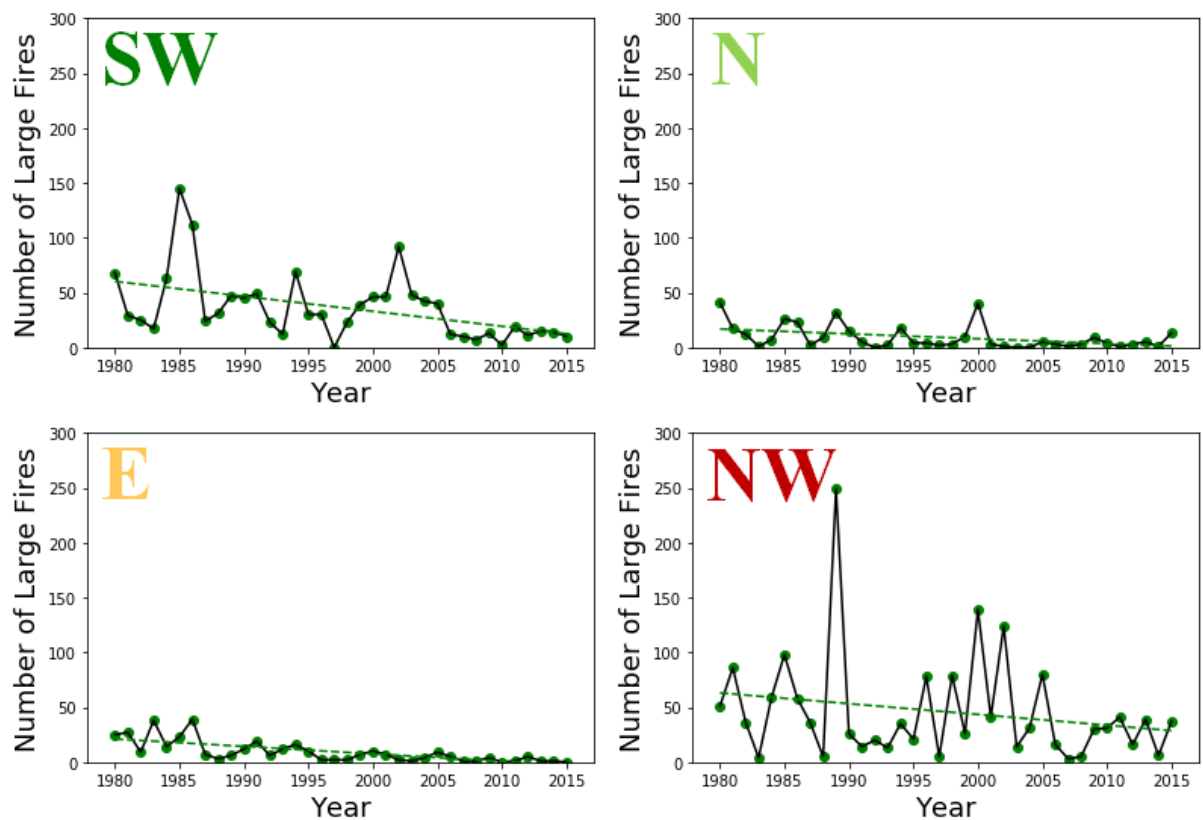


Figure A.2 – Trends of FWT_2 annual number of large fires (Mann- Kendall Test $p < 0.001$) for the SW and NW with a negative trend and N and E regions with no trend.

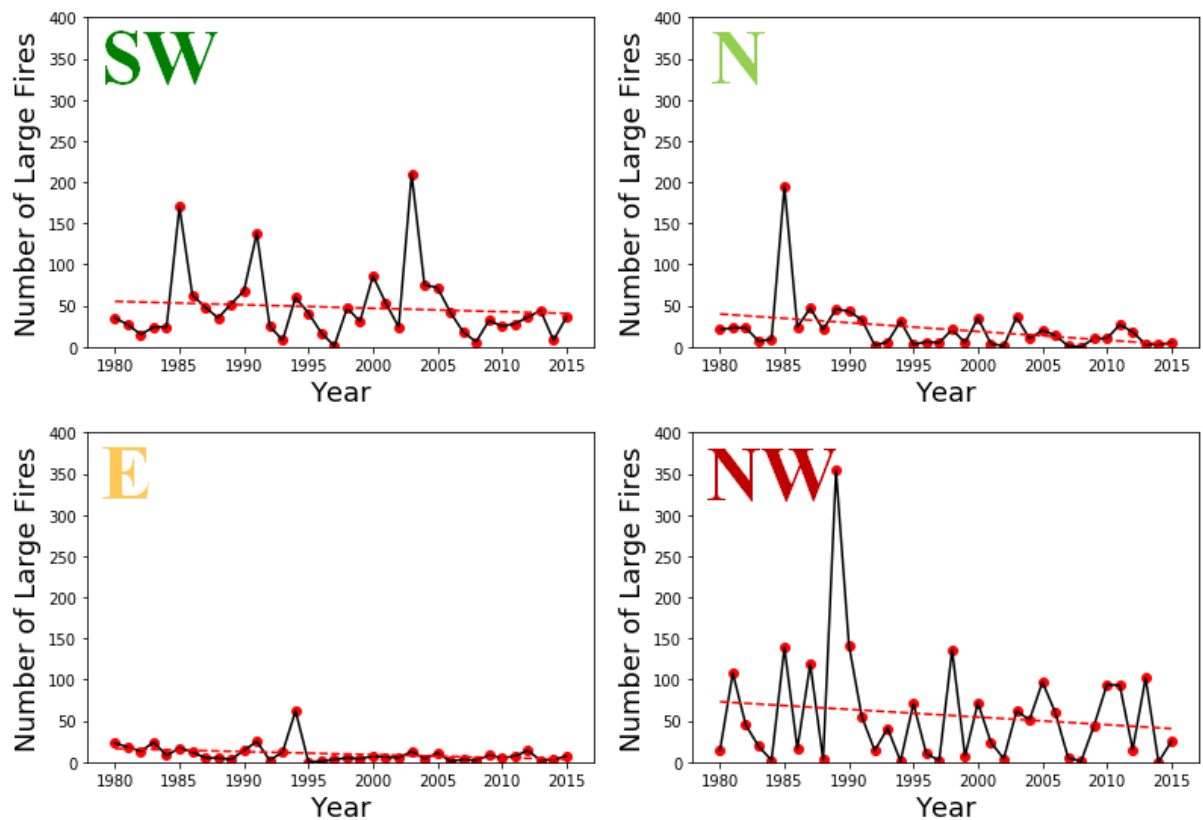


Figure A.3 - Trends of FWT_3 annual number of large fires (Mann- Kendall Test $p < 0.001$) for the N and E with a negative trend and NW and SW regions with no trend.